Forward.

These are step-by-verifiable-step notes designed to take students with a year of calculus based physics who are about to enroll in ordinary differential equations all the way to doctoral foundations in either mathematics and physics without mystery. Abstract algebra, topology (local and global) folds into a useful, intuitive toolset for ordinary differential equations and partial differential equations, be they linear or nonlinear. The algebraist, the topologist, the theoretical physicist, the applied mathematician and experimental physicist are artificial distinctions at the core. There is unity.

Mathematician, you will see step-by-verifiable-step algebra, topology (local and global) in a unified framework to treat differential equations, ordinary, partial, linear and nonlinear. You will then see why the physicists created a great font of differential equations, the calculus of variations. You will see why the physicists care about both discrete and continuous (topological) Lie groups and understand what quantum mechanics is as a mathematical system from its various historical classical physical roots: Lagrangian mechanics, Hamiltonian mechanics, Poisson brackets. You will have the tools to understand the Standard Model of physics and some of our main paths forward to grand unified theories and theories of everything. With these notes you should never again be able to practice abstraction for the sake of abstraction. Physicist, you will not be held hostage to verbiage and symbology. You will see that mathematics has deep, unavoidable limitations that underlie physics, itself suffering unavoidable limitations. You will see unity, *e.g.*, summing angular momentum in terms of tensor products and directions sums, ladder operators, Young's tableaux, root and weigh diagrams as different codifications of the same thing. Neither of you have to take your required courses as exercises in botany and voodoo as exemplified by ordinary differential equations. You will have context and operational skills. As lagniappes you will have the calculus of variations, the fractional calculus, stochastic calculus and stochastic differential equations.

Contents

Part I. (p. 1) Assuming only a mathematical background up to a sophomore level course in ordinary differential equations, Part I treats the application of symmetry methods for differential equations, be they linear, nonlinear, ordinary or partial. The upshot is the development of a naturally arising, systematic abstract algebraic toolset for solving differential equations that simultaneously binds abstract algebra to differential equations, giving them mutual context and unity. In terms of a semester of study, this material would best follow a semester of ordinary differential equations. The algorithmics, which will be developed step by step with plenty of good examples proceed along as follows: (1) learn to use the linearized symmetry condition to determine the Lie point symmetries, (2) calculate the commutators of the basis generators and hence the sequence of derived subalgebras, (3) find a sufficiently large solvable subalgebra, choose a canonical basis, calculate the fundamental differential invariants, and (4) rewrite the differential equation in terms of any differential invariants; then use each generator in turn to carry out one integration. This sounds like a mouthful, but you will see that it is not. The material is drawn from my notes derived from "Symmetry Methods for Differential Equations: A Beginner's Guide," Peter E. Hydon, Cambridge University Press, 2000.

Part II. (p. 125) Part II which assumes no additional background, should be learned in parallel with Part I. Part II builds up the calculus of variations by paralleling the buildup of undergraduate elementary calculus. Present day physics including classical mechanics, electrodynamics, quantum physics, quantum field theories, general relativity, string theories, and loop quantum gravity, for example, are all expressed in terms of some variational principle extremizing some action. This extremization process leads to, through the calculus of variations, sets of differential equations. These differential equations have associated symmetries (Part I) that underlie our present understanding of fundamental physics, the Standard Model. The same goes for many of the theoretical symmetries stretching beyond the Standard Model such as grand unified theories (GUTs) with or without supersymmetry (SUSY), and theories of everything (TOEs) like string theories. We will buttress the use of this variational toolset with history. Part II will thus provide us a deep, practical, intuitive source of differential equations, while Part I places the investigation of these differential equations in a general, practical, intuitive, algebraic framework readily accessible to the college sophomore. The variational material is drawn from "Calculus of Variations (Dover Books on Mathematics)," Lev D. Elsgolc, Dover Publications, 2007.

Beyond the sophomore level material contained in Parts I and II, a student pursuing deeper studies in either physics or mathematics will have already been well served by these notes, forever understanding that most abstract mathematics is likely clothed within a rich, intuitively unified, and useful context, and that any voodoo mathematical prescriptions in physics can be deconstructed from a relatively small collection of fundamental mathematical tools and physical principles. It would be desirable, but not necessary for studying Parts I and II, to have had some junior level exposure to classical mechanics motivating the calculus of variations. In lieu of this, I strongly recommend parallel readings from "Variational Principles In Dynamics and Quantum Theory (Dover Books on Physics)", Wolfgang Yourgrau and Stanley Mandelstam, Dover Publications, 1979 to tie the calculus of variations to mechanics, quantum mechanics and beyond through the historical development of this field from Fermat to Feynman.

Part III. (p. 136) Part III culminates the unifying goal of Parts I and II at the junior level, intuitively unifying algebra and topology together into algebraic topology with applications to differential equations and physics. Whereas in Part I students learn to study the sub-algebraic structure of the commutators of a differential equation to learn if a given differential equation may be more readily solved in new coordinates and/or reduced in order, Part III begins to develop the topological linkage of commutators to quantum field theories and general relativity by intuitively developing the concepts underlying parallel transport and the covariant derivative. This picture is developed step-by-step free of hand waving. It is recommended that the material in Part III be studied in parallel with a traditional junior level course in partial differential equations. If you are doing Part III solo, the parallel material for partial differential equations can be found in, "Applied Partial Differential Equations with Fourier Series and Boundary Value Problems," 4th ed., Richard Haberman, Prentice Hall, 2003—a great text in any edition. Having a junior level background in classical mechanics up to the Lagrangian and Hamiltonian approaches would greatly add to the appreciation of Parts I, II, and III. In lieu of this background is a reading of the previously cited Dover history book by Wolfgang Yourgrau and Stanley Mandelstam. The material covering algebraic topology and differential equations is drawn from the first four chapters of "Lie Groups, Lie Algebras, and Some of Their Applications (Dover Books on Mathematics)", Robert Gilmore, Dover Publications, 2006. Part III cleans up and fills in Gilmore's Dover book. The nitty gritty connections to quantum field theories and general relativity can be further elaborated from reading chapter three of "Quantum Field Theory," 2nd ed., Lewis H. Ryder, Cambridge University Press, 1996, and from "A Short Course in General Relativity," 3rd ed., James A. Foster and J. David Nightingale, Springer, 2005. The latter book provides a step by step construction of general relativity including the concept of parallel transport and the covariant derivative from a geometric point of view.

Part IV. (p. 207) In Part IV machinery is built up to study the group theoretic structures associated, not with differential equations this time, but with polynomials to prove the insolvability of the quintic, now providing the student with two examples of the utility of abstract algebra and linear algebra to problems in mathematics. As we develop this machinery, we will study the properties of the irreducible representations of discrete groups with applications to crystallography, molecular vibrations, and quantum physics. The toolset is then extended to continuous groups with applications to quantum physics and particle physics. It will be shown how to use ladder algebras to solve quantum mechanical differential equations algebraically. No step will be left out to mystify students. The material dealing with the insolvability of the quintic and discrete groups is pulled from two sources, Chapter 1 of "Lie Groups, Physics, and Geometry: An Introduction for Physicists, Engineers and Chemists," Robert Gilmore, Cambridge University Press, 2008, and from the first four chapters of the first edition of "Groups, Representations and Physics," 2nd ed., H. F. Jones, Institute of Physics Publishing, 1998. The latter, 2008 book by Robert Gilmore is too fast paced, and too filled with hand waving to serve other than as a guide for what is important to learn after learning the material presented in this work. The material dealing with continuous groups is derived from many sources which I put together into a set of notes to better understand "An Exceptionally Simple Theory of Everything," A. Garrett Lisi, arXiv:0711.0770v1, 6 November 2007.

A note on freebies. It is assumed that the student will take a course in complex analysis at the level of any late edition of "Complex Variables and Applications," James W. Brown and Ruel V. Churchill, McGraw-Hill Science/Engineering/Math, 8th ed., 2008. Strongly recommended are one or two good courses in linear algebra. A physics student, typically at the graduate level, is usually required to take a semester of mathematical physics covering a review of undergraduate mathematics and a treatment of special functions and their associated, physics-based differential equations. Once again the student is back to studying botany, and once again symmetry groups unify the botany. A free book treating this can be downloaded from http://www.ima.umn.edu/~miller/lietheoryspecialfunctions.html ("Lie Theory and Special Functions," by Willard Miller, Academic Press, New York, 1968 (out of print)).

A note on step-by-step books. "Introduction to Electrodynamics," 3rd ed., David J. Griffiths, Benjamin Cummings, 1999, or equivalent level of junior level electrodynamics, "Quantum Electrodynamics," 3rd ed., Greiner and Reinhardt, Springer 2002, and "Quantum Field Theory," 2nd ed., Lewis H. Ryder, Cambridge University Press, 1996 are each excellent, clearly written and self-contained.

Griffiths should be read from end to end. Greiner and Reinhardt should be read up to at least chapter four, if not up to chapter five to gain hands-on experience with calculating cross sections and decay rates the old fashioned way that led Richard Feynman to develop the Feynman diagram approach. The introductory chapter of Ryder may be skipped. Chapters two and three are where the physics and the mathematics lay that are relevant to much of the material presented in this work. The introductory chapter on path integrals is also pertinent.

Part V. (p. 294) Part V treats a miscellany of topics. Principal among these topics is material drawn from "The Fractional Calculus, Theory and Applications of Differentiation and Integration to Arbitrary Order (Dover Books on Mathematics)" by Keith B. Oldham and Jerome Spanier, Dover Publications, 2006. You will not only come to appreciate the gamma function better, you will be able to ask if the $\sqrt{i\pi}^{\text{th}}$ derivative has any meaning. Who said we can only take 1^{st}, 2^{nd},…,n^{th} order derivatives, or integrate once, twice,…,or *n* times? Calculus is more general, more unified, more intuitive, and more physical than calculus with only integer order differentiation or integration. The conversation will then turn to stochastic processes (p. 301). In my studies, I found measure theoretic analysis to be another source of meaningless, isolated, dry crap until I got into financial physics and needed to work with stochastic differential equations. Finance and statistical physics, to a lesser extent as currently taught in graduate physics courses, give context to measure theory. Material to show this is taken from "Options, Futures, and Other Derivatives," 5^{th} (or higher) edition, John C. Hull, Prentice Hall, 2002, as well as from personal notes.

Here is a final word to the physicists. Parts I and II of this work go to support graduate level classical mechanics and its extensions to quantum physics and quantum field theory. You should buttress your understanding of classical mechanics beyond the standard graduate course covering the material in, say, "Classical Mechanics," 3^{rd} ed., Herbert Goldstein, Addison Wesley, 2001. Goldstein certainly provides a good treatment of classical mechanics, giving the reader the background underlying the development of quantum physics, but he does not cover continuum mechanics. Continuum mechanics is not just for the engineer. The development of tensors and dyadic tensors is far greater in continuum mechanics than in typical, introductory general relativity. It is your loss not to acquire this more general toolset. I recommend, "Continuum Mechanics (Dover Books on Physics)," A. J. M. Spencer, Dover Publications, 2004. To complete one's understanding of mechanics, one should also study "Exploring Complexity," G. Nicolis and I. Prigogine, W H Freeman, 1989. Moving from complexity to statistical physics, my favorite statistical physics book is "A Modern Course in Statistical Physics," Linda E. Reichl, Wiley-VCH, 2009 (or its older edition). For experience solving practical problems, study "Statistical Mechanics (North-Holland Personal Library)", R. Kubo, H. Ichimura, T. Usui, and N. Hashitsume, North Holland, 1990. Underlying statistical physics is thermodynamics. I recommend "thermodynamics (Dover Books on Physics)" by Enrico Fermi, Dover Publications, 1956. "An Introduction to equations of state: theory and applications" by S. Eliezer, A. G. Ghatak and H. Hora (1986) gives a pretty good treatment of where our knowledge in thermodynamics and statistical physics abuts our ignorance, as well as shows how quickly mathematical models and methods become complex, difficult and approximate. This compact book has applications far outside of weapons work to work in astrophysics and cosmology. Rounding out some of the deeper meaning behind statistical physics is information theory, I recommend reading "The Mathematical Theory of Communication" by Claude E. Shannon and Warren Weaver, University of Illinois Press, 1998, and "An Introduction to Information Theory, Symbols, Signals and Noise" by John R. Pierce (also from Bell Labs), Dover Publications Inc., 1980. The material covered in Part V on measure theory and stochastic differential equations fits well with the study of statistical physics.

Parts I and II also go to the study of electrodynamics, undergraduate and graduate. At the graduate level ("Classical Electrodynamics," 3rd ed., John D. Jackson, Wiley, 1998) one is inundated with differential equations and their associated special functions. The online text by Willard Miller tying special functions to Lie symmetries is very useful at this point. I also recommend an old, recently reprinted book, "A Course of Modern Analysis," E. T. Whittaker, Book Jungle, 2009. The historical citations, spanning centuries, are exhaustive.

Parts I through V of this work underlie studies in general relativity, quantum mechanics, quantum electrodynamics and other quantum field theories. With this background you will have better luck reading books like, "A First Course in Loop Quantum Gravity," R. Gambini and J. Pullin, Oxford University Press, 2011, and "A First Course in String Theory," Barton Zweibach, 2nd ed., Cambridge University Press, 2009. Again, only together do mathematics and physics provide us with a general, intuitive grammar and powerful, readily accessible tools to better understand and explore nature and mathematics, and even to help us dream and leap beyond current physics and mathematics. Before you get deep into particle physics, I recommend starting with "Introduction to Elementary Particles," 2nd ed., David Griffiths, Wiley-VCH, 2008.

A. Alaniz

Apologies for typos in this first edition, December 2012. Teaching should be more than about how, but also about why and what for. Read this stuff in parallel, in series, and check out other sources. Above all, practice problems. The file "Syllabus" is a words based syllabus for the mathematician and physicist, and a recounting of the origin of some of the main limitations of mathematics and physics.

Part I. **Chapter 1**. (Accessible to sophomores; required for mathematics/physics majors up through the postdoctoral research level) We begin with an example.

Example 1.1—The symmetry of an ordinary differential equation (ODE). The general solution of

$$\frac{dy}{dx} = \frac{2y}{x} \tag{1}$$

is $y = cx^2$. We restrict our attention to $x > 0, y > 0$, in which each solution curve corresponds to a particular $c > 0$. The set of solution curves is mapped to itself by the discrete symmetry

$$(\hat{x}, \hat{y}) = \left(\frac{x}{y}, \frac{1}{y}\right). \tag{2}$$

If we pick one particular solution curve of (1), say $c = c_1$, then

$$(\hat{x}, \hat{y}) = \left(\frac{x}{y}, \frac{1}{y}\right) = \left(\frac{x}{c_1 x^2}, \frac{1}{c_1 x^2}\right) = \left(\frac{1}{c_1 x}, \frac{1}{c_1 x^2}\right). \tag{3}$$

Solving for x gives us $x = 1/(c_1 \hat{x}\ )$. Then $\hat{y} = 1/(c_1 x^2) = {c_1}^2 \hat{x}^2 / c_1 = c_1 \hat{x}^2$. Symmetry (2) to the solution of ODE (1) is a "symmetry" because it leaves the form of the solution invariant in either the (x, y) coordinates or the $(\hat{x}, \hat{y})$ coordinates, like rotating a square by ninety degrees on the plane leaves the square invariant. The symmetry is a smooth (differentiable to all orders) invertible transformation mapping solutions of the ODE to solutions of the $\widehat{\text{ODE}}$. Invertible means the Jacobian is nonzero:

$$\hat{x}_x \hat{y}_y - \hat{x}_y \hat{y}_x \neq 0. \tag{4}$$

Another way to express the transformation is using a matrix:

$$\begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} = \begin{pmatrix} \frac{1}{c_1 x^2} & 0 \\ 0 & 1/y^2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{c_1 x^2} & 0 \\ 0 & 1/(c_1^2 x^4) \end{pmatrix} \begin{pmatrix} x \\ c_1 x^2 \end{pmatrix}. \tag{5}$$

Since we restricted ourselves to $x > 0, y > 0$, it must be that $c > 0$. The inverse of the matrix is $\begin{pmatrix} c_1 x^2 & 0 \\ 0 & y^2 \end{pmatrix}$, and it is also smooth. In these notes, if x is any point of the object (a point on a solution curve of the ODE in our case), and if $\Gamma: x \longmapsto \hat{x}(x)$ is a symmetry, then we assume $\hat{x}$ to be infinitely differentiable wrt x. Since Γ^{-1} is also a symmetry, then x is infinitely differentiable wrt $\hat{x}$. Thus Γ is a (C^{∞}) diffeomorphism (a smooth invertible mapping whose inverse is also smooth). Is the connection between symmetry (an algebraic concept) and a differential equation deep or merely superficial?

Example 1.2—(More evidence tying symmetries to differential equations) Consider the Riccati equation

$$y' = xy^2 - \frac{2y}{x} - \frac{1}{x^3} = \omega(x,y), \qquad x \neq 0. \tag{6}$$

Let's consider a one-parameter symmetry more focused on the ODE than its solution. Let $(\hat{x},\hat{y}) = (e^{\varepsilon}x, e^{-2\varepsilon}y)$. Then $(x,y) = (e^{-\varepsilon}\hat{x}, e^{2\varepsilon}\hat{y})$. Substituting into (6) to get

$$\frac{dy}{dx} = e^{3\varepsilon}\frac{d\hat{y}}{d\hat{x}} = e^{3\varepsilon}\hat{x}\hat{y}^2 - e^{3\varepsilon}\frac{2\hat{y}}{\hat{x}} - e^{3\varepsilon}\frac{1}{\hat{x}^3}, \qquad \hat{x} \neq 0. \tag{7}$$

$$\hat{y}' = \hat{x}\hat{y}^2 - \frac{2\hat{y}}{\hat{x}} - \frac{1}{\hat{x}^3}, \qquad \hat{x} \neq 0. \tag{8}$$

How did we cook up this symmetry? By taking a guess, an ansatz. A more satisfying answer lays ahead.

If we set ε to zero, the symmetry is the identity symmetry. As we vary the parameter ε we trace a curve in the $\hat{x}\hat{y} -$ plane. At any given (x,y) the tangent to the curve parameterized by ε is

$$\left(\left(\frac{d\hat{x}}{d\varepsilon}\right)_{\varepsilon=0}, \left(\frac{d\hat{y}}{d\varepsilon}\right)_{\varepsilon=0}\right) = \big(\xi(x,y), \eta(x,y)\big) = (x, -2y). \tag{9}$$

If we suppose that the tangent to the curve parameterized by ε is parallel to y', then we are supposing that

$$y' = \frac{\eta(x,y)}{\xi(x,y)} = \omega(x,y). \tag{10}$$

Then,

$$\eta(x,y) - \omega(x,y)\xi(x,y) = 0. \tag{11}$$

Equivalently,

$$-2y - \left(xy^2 - \frac{2y}{x} - \frac{1}{x^3}\right)x = \frac{1}{x^2} - x^2y^2 = 0. \tag{12}$$

This is true if $y = \pm 1/x^2$. So we have found some solutions! Let's check if $y = -1/x^2$ is a solution.

$$\frac{dy}{dx} = \frac{2}{x^3} = xy^2 - \frac{2y}{x} - \frac{1}{x^3} = \frac{x}{x^4} + \frac{x}{x^4} - \frac{1}{x^3} = \frac{2}{x^3}. \tag{13}$$

It seems that if we can find a symmetry to a differential equation we might find some solutions. This was the great observation of Sophus Lie. There are other solutions.

Let's build some tools. We restrict ourselves to first order ODEs. Soon afterwards we shall build tools for higher order ODEs and partial differential equations (PDEs), either linear or nonlinear ODEs or PDEs. Consider the first order differential equation

$$\frac{dy}{dx} = \omega(x, y). \tag{14}$$

We assume there is a diffeomorphism $\Gamma\colon (\mathrm{x}, \mathrm{y}) \longmapsto (\hat{\mathrm{x}}, \hat{\mathrm{y}})$ that is also a symmetry of ODE (14). That is we assume that

$$\frac{d\hat{y}}{d\hat{x}} = \omega(\hat{x}, \hat{y}) \ \text{ when } \ \frac{dy}{dx} = \omega(x, y). \tag{15}$$

Equation (15) is called the "symmetry condition" for ODE (14). The symmetry condition is a symmetry transformation (like the one in example 1.2) which leaves the differential equation invariant despite the smooth change of coordinates to $(\hat{x}, \hat{y})$.

Does Γ exist? I don't know. The point is to assume at least one such symmetry exists that is a diffeomorphism connecting $(\hat{x}, \hat{y})$ coordinates to (x, y) coordinates, *i.e.*, $\hat{x} = \hat{x}(x, y)$ and $\hat{y} = \hat{y}(x, y)$, and to study the properties that such symmetry must have as a consequence of our smoothness stipulations.

Relating $d\hat{y}/d\hat{x}$ to the original coordinates (x, y) is the total derivative $D_x = \partial_x + y'\partial_y + y''\partial_{y'} + \cdots$. In this notation, subscript Latin letters imply differentiation wrt that Latin letter, *e.g.*, $\hat{y}_y = \partial\hat{y}/\partial\, y$. Keeping only terms up to first order

$$\frac{D_x\hat{y}}{D_x\hat{x}} = \frac{\hat{y}_x + y'\hat{y}_y}{\hat{x}_x + y'\hat{x}_y}. \quad (16)$$

The symmetry condition (15) for ODE (14) yields

$$\frac{D_x\hat{y}}{D_x\hat{x}} = \frac{\hat{y}_x + y'\hat{y}_y}{\hat{x}_x + y'\hat{x}_y} = \omega(\hat{x}, \hat{y}) \text{ when } \frac{dy}{dx} = \omega(x, y). \quad (17)$$

Since $y' = \omega(x, y)$ in the original coordinates, and we may write

$$\frac{\hat{y}_x + \omega(x, y)\hat{y}_y}{\hat{x}_x + \omega(x, y)\hat{x}_y} = \omega(\hat{x}, \hat{y}). \quad (18)$$

Equation (18) together with the requirement that Γ is a diffeomorphism is equivalent to the symmetry condition (15). Equations (17) or (18) tie (or "ligate") the original coordinates (x, y) to the new coordinates $\left(\hat{x}(x, y), \hat{y}(x, y)\right)$. This result is important because it may lead us to some if not all of the symmetries of an ODE. Was the symmetry to the Riccati equation pulled from the ass of some genius, or was there a method to the madness? (Notice the Riccati equation is nonlinear.)

Example 1.3—To better understand the hunt for symmetries, consider the simple ODE

$$\frac{dy}{dx} = y. \quad (19)$$

Symmetry condition (17) implies that each symmetry of (19) satisfies the PDE

$$\frac{\hat{y}_x + y'\hat{y}_y}{\hat{x}_x + y'\hat{x}_y} = \hat{y}.$$

Since $y' = y$ in the original (x, y)coordinates, equation (18) equivalently implies that

$$\frac{\hat{y}_x + y\hat{y}_y}{\hat{x}_x + y\hat{x}_y} = \hat{y}. \tag{20}$$

Rather than trying to find the general solution to this PDE, let us instead use (20) to inspire some simple guesses at some possible symmetries to ODE (19). Say we try $(\hat{x}, \hat{y}) = (\hat{x}(x, y), y)$. That is $\hat{y}$ is only a function of y, and not of x, *i.e.*, $\hat{y} = y$. Then $\hat{y}_x = 0$, $\hat{y}_y = 1$, and (20) reduces to

$$\frac{y}{\hat{x}_x + y'\hat{x}_y} = y, \tag{21}$$

or,

$$\hat{x}_x + y'\hat{x}_y = 1, \tag{22}$$

For any symmetries which are diffeomorphisms, the Jacobian is nonzero. That is,

$$\hat{x}_x\hat{y}_y - \hat{x}_y\hat{y}_x = \hat{x}_x \cdot 1 - \hat{x}_y \cdot 0 = \hat{x}_x \neq 0. \tag{23}$$

The simplest case of $\hat{x}_x \neq 0$ is $\hat{x} = x + \varepsilon$. So the simplest one-parameter symmetry to ODE (19) is

$$(\hat{x}, \hat{y}) = (x + \varepsilon, y). \tag{24}$$

Let's check by substituting this into the LHS and RHS of (17).

$$\left\{\frac{\hat{y}_x + y'\hat{y}_y}{\hat{x}_x + y'\hat{x}_y} = \frac{0 + y' \cdot 1}{1 + 0} = y'\right\}_{\text{LHS}} = \{\hat{y} = y\}_{\text{RHS}}. \tag{25}$$

The diffeomorphism $(\hat{x}, \hat{y}) = (x + \varepsilon, y)$ is therefore a symmetry of ODE (19). We now have some hope that producing the symmetry to the Riccati equation may have more to it than a genius' guess. There are more powerful, more systematic methods to come.

WARNING!!! Eleven double spaced pages with figures and examples follow before we treat the Riccati equation with a more complete set of tools (feel free to peek at example 1.8). Most of the material is fairly transparent on first reading, but some of it will require looking ahead to the fuller Riccati example (example 1.8), going back and forth through these notes.

Let's collect a lot of equivalent verbiage and nomenclature. In example 1.3, symmetry (24) to ODE (19) is called a one-parameter symmetry because it (a) leaves the form of the given ODE invariant in either the original coordinates (x, y), or in the new coordinates $(\hat{x}, \hat{y})$, just as rotating a square by ninety degrees on the plane leaves it invariant, and because (b) the symmetry depends smoothly on one real number parameter, ε. Symmetry (24) is a smooth mapping (diffeomorphism). Its Jacobian is nonzero.

Symmetry (24) may also be expressed in matrix form as

$$\begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} = \begin{pmatrix} 1 & \varepsilon/y \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + \varepsilon \\ y \end{pmatrix}, \quad y \neq 0. \tag{26}$$

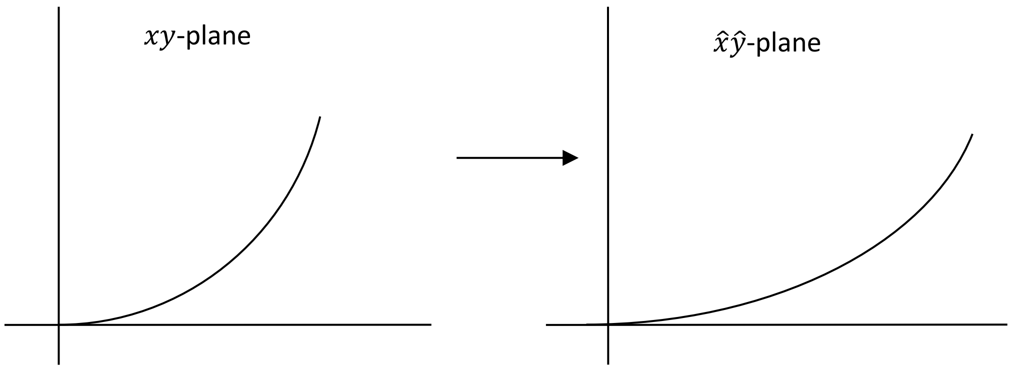
Matrix (26) is invertible. I'm telling you this matrix stuff because we shall eventually see that there is practical value to studying the abstract (group theoretic) algebraic properties of the matrix "representations" of the symmetries of a differential equation, as well as to studying the topological properties of such symmetries, like their continuity and compactness. Abstract algebra, topology, and algebraic topology aren't vacuous constructs built for useless mental masturbation by "pure" mathematicians; the study of differential equations is more than a study of botany. Anytime during these eleven pages that you feel discouraged, please peek ahead to example 1.8 to see that it's worth it.

In example 1.1, we found one symmetry to ODE (1), namely symmetry (2). ODE (1) has another symmetry, namely,

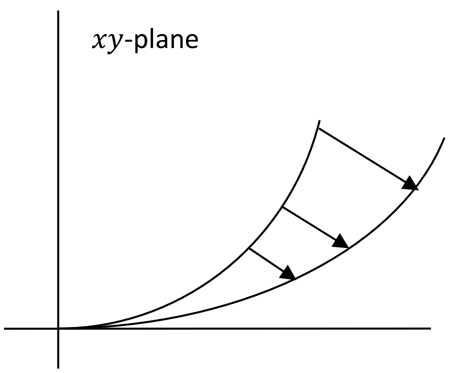
$$(\hat{x}, \hat{y}) = (e^{\varepsilon} x, e^{-\varepsilon} y). \tag{27}$$

Check by substitution that (27) is a one-parameter symmetry of ODE (1). So there can be more than one symmetry.

The solution curve $y = c_1 x^2$ gets mapped to $(\hat{x}, \hat{y}) = (e^{\varepsilon} x, c_1 e^{-\varepsilon} x^2)$. Solving for x, we obtain $x = e^{-\varepsilon}\hat{x}$, and therefore $\hat{y} = c_1 e^{-3\varepsilon} \hat{x}^2 = c_2 \hat{x}^2$, the same form as in (x, y) coordinates. We can see that the $\hat{x}\hat{y}$-plane and the xy-plane contain the same set of solution curves.



There is another point of view. Instead of a transformation from one plane to another, we can also imagine that the symmetry "acts" on a solution curve in the xy-plane as depicted in the following figure.



In the latter point of view, the symmetry is regarded as a mapping of the *xy*-plane to itself, called the *action* of the symmetry on the *xy*-plane. Specifically, the point with the coordinates (x, y) is mapped to the point whose coordinates are $(\hat{x}, \hat{y}) = (\hat{x}(x, y), \hat{y}(x, y))$. The solution curve $y = f(x)$ is the set of points with coordinates$(x, f(x))$. The solution curve is mapped to $(\hat{x}, \tilde{f}(\hat{x}))$ by the symmetry, *e.g.*, $(x, c_1x^2) \mapsto (\hat{x}, c_2\hat{x}^2)$. The solution curve is ***invariant*** under the symmetry if $f = \tilde{f}$. A symmetry is ***trivial*** if it leaves every solution curve invariant. Symmetry (2) to ODE (1) is trivial. Go back and see for yourself. Symmetry (27) to ODE (1) is not trivial.

A one-parameter symmetry depends on only one-parameter, *e.g.*, ε. A one-parameter symmetry may look like $(\hat{x}, \hat{y}) = (e^{\varepsilon}x, e^{-\varepsilon}y)$ or like $(\hat{x}, \hat{y}) = (x + \varepsilon, y + \varepsilon)$. A two-parameter symmetry may look like $(x + \varepsilon_1, y + \varepsilon_2)$. We restrict ourselves to one-parameter symmetries until further notice.

Group theory basics. It is time to note that our one-parameter symmetries are groups in the sense of modern algebra. Why? To masturbate with nomenclature as you do in an abstract algebra class? No. Because, as you will soon see, studying the group structure of a symmetry of a differential equation will have direct relevance to reducing its order to lower order, and will have direct relevance to finding some, possibly all of the solutions to the given differential equation—ordinary, partial, linear, or nonlinear. So what is a group?

A ***group*** is a set G together with a binary operation * such that

(I) There is an element of G, called the identity (e), such that if $g \in G$, the $e * g = g * e = g$. (**Identity**)

From ODE (1) with symmetry $(\hat{x}, \hat{y}) = (e^{\varepsilon}x, e^{-\varepsilon}y)$, we see that $(\hat{x}, \hat{y}) = (e^{\varepsilon}x, e^{-\varepsilon}y) = (x, y)$ when $\varepsilon = 0$. The identity element doesn't do jack. It has no action.

(II) For every ($\forall$) $a \in G$ there exists ($\exists$) a $a^{-1} \in G$ such that ($\ni$) $a * a^{-1} = a^{-1} * a = e$. (**Inverse**)

The inverse to the symmetry for ODE (1) $(\hat{x}, \hat{y}) = (e^{\varepsilon}x, e^{-\varepsilon}y)$ is $(e^{-\varepsilon}x, e^{\varepsilon}y)$. Actually, these two symmetries are their own inverses. The first symmetry "moves" (x, y) to $(e^{\varepsilon}x, e^{-\varepsilon}y)$. Applying the inverse moves you back to $(e^{\varepsilon-\varepsilon}x, e^{-\varepsilon+\varepsilon}y) = (x, y)$. In the symmetry we are using, you could reverse the order and get the same result, but this is not always the case. Not everything is commutative (Abelian). Cross products of three dimensional Cartesian vectors, for example, are not commutative (non Abelian). $\mathbf{A} \times \mathbf{B} \neq \mathbf{B} \times \mathbf{A}$. (Titillation: Noncommutativity underlies quantum physics. You will see this before you finish this set of notes.)

(III) If $a, b \in G$, then $a * b \in G$. (**Closure**)

If a denotes $(e^{\varepsilon_1}x, e^{-\varepsilon_1}y)$ and b denotes $(e^{\varepsilon_2}x, e^{-\varepsilon_2}y)$, then $a * b$ denotes $(e^{\varepsilon_1+\varepsilon_2}x, e^{-\varepsilon_1-\varepsilon_2}y)$, which is a member of G. Note that since ε_1, the group G is a ***continuous group***.

(IV) If $a, b, c \in G$, then $a * (b * c) = (a * b) * c \in G$. (**Associativity**)

Check this yourself as we checked property (III). Lastly note that since the parameters are continuous, the groups are *continuous groups*. Lots of diverse mathematical structures are groups.

Example 1.4—The set of even integers together with addition being * is a ***discrete group***. Check the definition.

Example 1.5—The set of polynomials together with the rules of polynomial addition being * is a group.

Example 1.6—The set of 2×2matrices that are invertible (have inverses) form a group with matrix multiplication being *. The identity is the 2×2 matrix with ones in the diagonal and zeros otherwise. That is, $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. (Hint: In linear algebra we learn about matrices that are commutative, and matrices that are not commutative—this stuff underlies quantum physics.)

The symmetry (27) to ODE (1) is, moreover, an infinite one-parameter group because the parameter ε is a real number on the real number line. In example 1.3, we met an infinite set of continuously connected symmetries, namely, $(\hat{x}, \hat{y}) = (x + \varepsilon, y)$. This symmetry smoothly maps the plane $\mathbb{R}^2$ to the plane $\mathbb{R}^2$. Later on, with ODEs of higher than first order, we shall deal with symmetries from $\mathbb{R}^3$ to $\mathbb{R}^3$, from $\mathbb{R}^4$ to $\mathbb{R}^4$, and so on from $\mathbb{R}^n$ to $\mathbb{R}^n$. These symmetries are called "Lie" groups after Sophus Lie, the dude who brought them to the fore to deal with differential equations. We deal with two more chunks of additional structure before getting back to practical, step-by-step applications to linear and nonlinear differential equations.

Theorem I. Let us suppose that the symmetries of $\frac{dy}{dx} = \omega(x, y)$ include the Lie group of translations $\omega(x, y) = \omega(x, y + \varepsilon)$ for all ε in some neighborhood of zero. Then,

$$\omega(x, y + \varepsilon) - \omega(x, y) = 0,$$

therefore when $\varepsilon \neq 0$ so does

$$\frac{\omega(x, y + \varepsilon) - \omega(x, y)}{\varepsilon} = 0,$$

and therefore

$$\lim_{\varepsilon \to 0} \frac{\omega(x, y + \varepsilon) - \omega(x, y)}{\varepsilon} \equiv \omega_y(x, y) = 0.$$

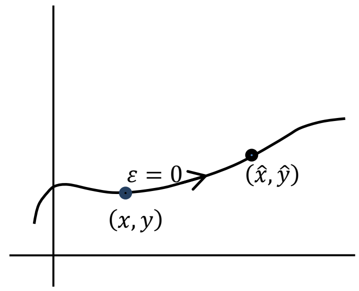
Thus the function ω only depends on x. Thus $\frac{dy}{dx} = \omega(x)$, and $y = \int \omega(x)dx + c$. The particular solution corresponding to $c = 0$ is mapped by the translation symmetry to $\hat{y} = \int \omega(x)dx + \varepsilon = \int \omega(\hat{x})d\hat{x} + \varepsilon$, which is the solution corresponding to $c = \varepsilon$. Note—A differential equation is considered solved if it has been reduced to quadrature—all that remains, that is, is to evaluate an integral. Also note that this theorem will be deeply tied to the use of canonical coordinates up ahead.

Action. It is useful to study the action of one-parameter symmetries on points of the plane. The ***orbit*** of a one-parameter Lie group through (x, y) is a set of points to which (x, y) can be mapped by a specific choice of ε,

$$(\hat{x}, \hat{y}) = \big(\hat{x}(x, y; \varepsilon), \hat{y}(x, y; \varepsilon)\big). \quad (28)$$

with initial condition

$$\big(\hat{x}(x, y; 0), \hat{y}(x, y; 0)\big) = (x, y). \quad (29)$$



The orbit through a point may be smooth as in the figure above, but there may be one or more invariant points. An ***invariant point*** is a point that gets mapped to itself by any Lie symmetry. An invariant point is a zero-dimensional orbit of the Lie group. In symmetry (24) the origin is mapped to the origin. The origin is an invariant point. Orbits themselves are closed. That is, the action of a Lie group maps each point on an orbit to a point on the orbit. Orbits are invariant under the action of a Lie group.

The arrow in the figure above depicts the tangent vector at (x, y). The tangent vector to the orbit at $(\hat{x}, \hat{y})$ is $\big(\xi(\hat{x}, \hat{y}), \eta(\hat{x}, \hat{y})\big)$, where

$$\frac{d\hat{x}}{d\varepsilon} = \xi(\hat{x}, \hat{y}), \qquad \frac{d\hat{y}}{d\varepsilon} = \eta(\hat{x}, \hat{y}).$$

The tangent vector at $(x, y)$is $\big(\xi(x, y), \eta(x, y)\big)$

$$\xi(\hat{x}, \hat{y}) = \left(\frac{d\hat{x}}{d\varepsilon}\right)_{\varepsilon=0}, \qquad \eta(x, y) = \left(\frac{d\hat{y}}{d\varepsilon}\right)_{\varepsilon=0}. \tag{31}$$

Therefore to first order in ε,

$$\hat{x} = x + \varepsilon\xi(x, y) + O(\varepsilon^2), \tag{32}$$

$$\hat{y} = y + \varepsilon\eta(x, y) + O(\varepsilon^2). \tag{33}$$

The set of tangent vectors is a smooth vector field. In example 1.3 with symmetry (24) we get

$$\xi(x, y) = \left(\frac{d\hat{x}}{d\varepsilon}\right)_{\varepsilon=0} = 1, \qquad \eta(x, y) = \left(\frac{d\hat{y}}{d\varepsilon}\right)_{\varepsilon=0} = 0.$$

Plugging the above into equations (32) and (33) we get $\hat{x} = x + \varepsilon$ and $\hat{y} = y$, which is symmetry (24). I was just checking that all of this stuff is consistent. Note that an invariant point is mapped to itself by every Lie symmetry. Thus for an invariant point, we have $\xi(x, y) = \eta(x, y) = 0$ from (32) and (33). It's alright if this stuff doesn't yet mean much. It will soon allow us to find, if possible, a change of coordinates that may allow a given differential equation to be reduced to a simpler form. The first application will be to finding more solutions to the Riccati equation in "canonical" coordinates.

Characteristic equation. This next block of structure follows directly from example 1.2. See the paragraph containing equations (9) through (12) in example 1.2. Any curve *C* is an invariant curve $y = f(x)$ if and only if (iff) the tangent to *C* at each (x, y) is parallel to the tangent vector $\big(\xi(x, y), \eta(x, y)\big)$. This is expressed by the *characteristic equation*

$$Q(x, y, y') = \eta(x, y) - y'\xi(x, y). \tag{34}$$

C is parallel to (ε, η) iff $Q(x, y, y') = 0$. on C, or

$$y' = \frac{\eta(x, y)}{\xi(x, y)}. \tag{35}$$

This implies that

$$\frac{dy}{dx} = \omega(x, y) \tag{36}$$

can have its invariant solution characterized by

$$\tilde{Q}\big(x, y, \omega(x, y)\big) = \eta(x, y) - y'\xi(x, y). \tag{37}$$

Equation (37) is the ***reduced characteristic***.

The new shorthand applied to the Riccati equation in example 1.2 updates that example to the following. We were given $y' = xy^2 - \frac{2y}{x} - \frac{1}{x^3} = \omega(x, y),\ x \neq 0$. The ansatz was $(\hat{x}, \hat{y}) = (e^{\epsilon}x, e^{-2\epsilon}y)$. The tangent vector is $(\hat{x}, \hat{y}) = (x, -2y)$. The reduced characteristic is $\tilde{Q}(x, y) =$

$\eta(x,y) - y'\xi(x,y) = -2y - \left(xy^2 - \frac{2y}{x} - \frac{1}{x^3}\right) = \frac{1}{x^2} - x^2y^2$. $\tilde{Q}(x,y) = 0$ iff $y = f(x) = \pm\frac{1}{x^2}$, the "invariant solutions". Most symmetry methods use the tangent vectors rather than the symmetries themselves to seek out "better" coordinates to find solutions to differential equations.

Canonical coordinates. We use canonical coordinates when the ODE has Lie symmetries equivalent to a translation. Symmetry (24) gives us an example of a symmetry to an ODE which is a translation $(\hat{x},\hat{y}) = (x, y+\varepsilon)$. The ODE is greatly simplified under a change of coordinates to canonical coordinates, *e.g.*, the Riccati equation $y' = xy^2 - \frac{2y}{x} - \frac{1}{x^3} = \omega(x,y),\ x \neq 0$ turns to $\frac{ds}{dr} = \frac{1}{r^2-1}$.

Given $(\hat{x},\hat{y}) = (x, y+\varepsilon)$ with tangent vector $\big(\varepsilon(x,y), \eta(x,y)\big) = (0,1)$, we seek coordinates $(r,s) = \big(r(x,y), s(x,y)\big)$ such that $(\hat{r},\hat{s}) = (r, s+\varepsilon)$. Then the tangent vector is $\left(\left(\frac{d\hat{r}}{d\varepsilon}\right)_{\varepsilon=0} = 0, \left(\frac{d\hat{s}}{d\varepsilon}\right)_{\varepsilon=0} = 1\right)$. Using $\frac{d\hat{x}}{d\varepsilon} = \xi(\hat{x},\hat{y})$, $\frac{d\hat{y}}{d\varepsilon} = \eta(\hat{x},\hat{y})$, and the chain rule

$$\frac{dr}{dx}\frac{dx}{d\varepsilon} + \frac{dr}{dy}\frac{dy}{d\varepsilon} = 0, \tag{38}$$

$$\frac{ds}{dx}\frac{dx}{d\varepsilon} + \frac{ds}{dy}\frac{dy}{d\varepsilon} = 1, \tag{39}$$

we get

$$\xi(x,y)r_x + \eta(x,y)r_y = 0, \tag{40}$$

$$\xi(x,y)s_x + \eta(x,y)s_y = 1. \tag{41}$$

By smoothness the Jacobian is not zero. That is,

$$r_x s_y - r_y s_x \neq 0. \tag{42}$$

Therefore a curve of constant r and a curve of constant s cross transversely. Any pair of functions $r(x,y), s(x,y)$ satisfying (40) through (42) is called a pair of *canonical coordinates*. The curve of constant r corresponds (locally) with the orbit through the point (x,y). The orbit is invariant under the

Lie group, so r is the ***invariant canonical coordinate***. Note that canonical coordinates cannot be defined at an invariant point because the determining equation for s, namely $\xi(x,y)s_x + \eta(x,y)s_y = 1$, has no solution if $\xi = \eta = 0$, but it is always possible to normalize the tangent vectors (at least locally). Also note that canonical coordinates defined by (40) and (41) are not unique. If (r,s) satisfy (40) and (41) so do $(\hat{r},\hat{s}) = \big(F(r), s + G(r)\big)$. (Thus, without proof there is a degeneracy condition which states $F'(r) \neq 0$, but there is still plenty of freedom left.)

Canonical coordinates can be obtained from (40) and (41) through the method of characteristics. In the theory of ODEs, the characteristic equation is

$$\frac{dx}{\varepsilon(x,y)} = \frac{dy}{\eta(x,y)} = ds. \tag{43}$$

This is a system of ODEs. Here follows a definition. A ***first integral*** of a given first-order ODE

$$\frac{dy}{dx} = f(x,y) \tag{44}$$

is a nonconstant function $\phi(x,y)$ whose value is constant on any solution $y = y(x)$ of the ODE. Therefore on any solution curve $y = y(x)$,

$$\frac{d\phi}{dx} = \frac{\partial\phi}{\partial x} + \frac{\partial\phi}{\partial y}\frac{dy}{dx} = 0, \text{or} \tag{45}$$

$$\frac{d\phi}{dx} = \frac{\partial\phi}{\partial x} + f(x,y)\frac{\partial\phi}{\partial y} = 0,\ \ \phi_y \neq 0. \tag{46}$$

The general solution is $\phi(x,y) = c$. Suppose that $\xi(x,y) \neq 0$ in equation (40). Then let's rearrange equation (40) as

$$r_x + \frac{\eta(x,y)}{\xi(x,y)} r_y = r_x + f(x,y) r_y = 0. \tag{47}$$

Comparing (46) with (40), we see that the r is a first integral of

$$\frac{dy}{dx} = \frac{\eta(x,y)}{\xi(x,y)}. \tag{48}$$

So $r = \phi(x,y)$ is found by solving (48). It is an invariant canonical coordinate. Sometimes we can determine a solution $s(x,y)$ by inspection, else we can use $r = r(x,y)$ to write y as a function of r and x. The coordinate $s(r,x)$ is obtained from (43) by quadrature:

$$s(r,x) = \left(\int \frac{dx}{\varepsilon\big(x, y(r,x)\big)}\right)_{r=r(x,y)}, \tag{49}$$

where r is being treated as a constant.

If $\varepsilon(x,y) = 0$ and $\eta(x,y) \neq 0$ then the canonical coordinates are

$$r = x, \qquad s(r,x) = \left(\int \frac{dy}{\eta(r,y)}\right)_{r=x}. \tag{50}$$

Example 1.7—Every ODE of the form $y' = F(y/x)$ admits the one-parameter Lie group of scalings $(\hat{x},\hat{y}) = (e^{\varepsilon}x, e^{\varepsilon}y)$. Consider $\frac{dy}{dx} = \frac{ky}{x}$ as a very simple example (of course we know $y = cx^k$ is the solution). If $x \neq 0$, the canonical coordinate r is $r = yx^{-k} = \text{constant}$. Then $s(x,y) = \int \frac{dx}{x} = \ln|x|$. Thus $(r,s) = \big(yx^{-k}, \ln|x|\big)$. At $x = 0$ we need a "new coordinate patch": $r = y^{-1}x^k$, $s = \frac{1}{k}\ln|y|$. So what? Finding canonical coordinates reduces ugly ODEs into simpler ODEs. We're steps away from this.

Recall that Lie symmetries of an ODE are nontrivial iff

$$\eta(x,y) \neq \omega(x,y)\xi(x,y). \tag{51}$$

If ODE (14), $\frac{dy}{dx} = \omega(x,y)$, has nontrivial Lie symmetries equivalent to a translation, it can be reduced to quadrature by rewriting it in terms of canonical coordinates as follows. Let

$$\frac{ds}{dr} = \frac{s_x + \omega(x,y)s_y}{r_x + \omega(x,y)r_y}. \tag{52}$$

The right hand side of (52) can be written as a function of r and s using the symmetry. For a general change of variables $(x, y) \mapsto (r, s)$, the transformed ODE (52) would be of the form

$$\frac{ds}{dr} = \Omega(r, s), \tag{53}$$

for some function Ω. However, since we assume (r, s) are canonical coordinates, the ODE is invariant under the group of translations in the s direction:

$$(\hat{r}, \hat{s}) = (r, s + \varepsilon). \tag{54}$$

Thus from theorem I we know that

$$\omega_s(r, s) = 0, \tag{55}$$

and therefore

$$\frac{ds}{dr} = \Omega(r). \tag{56}$$

The ODE has been reduced to quadrature, and the general solution to ODE (56) is

$$s(x, y) - \int \Omega(r)dr = c. \tag{57}$$

Therefore the general solution to ODE (14) is

$$s(x, y) - \int^{r(x,y)} \Omega(r)dr = c. \tag{58}$$

This is great, but of course we must first determine the canonical coordinates by solving

$$\frac{dy}{dx} = \frac{\eta(x, y)}{\xi(x, y)}. \tag{48}$$

<u>Example 1.8</u>—Let's finally compute the Riccati equation with both barrels using our updated toolset.

$$y' = xy^2 - \frac{2y}{x} - \frac{1}{x^3} = \omega(x, y), \qquad x \neq 0. \tag{6}$$

As we know, a symmetry of (6) is $(\hat{x}, \hat{y}) = (e^{\epsilon}x, e^{-2\epsilon}y)$. The corresponding tangent vector is $(\varepsilon, \eta) = (x, -2y)$. The reduced characteristic $\tilde{Q}\big(x, y, \omega(x, y)\big) = \eta(x, y) - \omega(x, y)\xi(x, y)$ is

$$\tilde{Q}\big(x, y, \omega(x, y)\big) = 2y - \left(xy^2 - \frac{2y}{x} - \frac{1}{x^3}\right)x = \frac{1}{x^2} - x^2y^2. \tag{59}$$

$\tilde{Q}\big(x, y, \omega(x, y)\big) = 0$ iff $y = \pm\frac{1}{x^2}$. We stopped here before. Now we use our symmetry's tangent vector to give us canonical coordinates to simplify the Riccati equation. Equation (48) becomes

$$\frac{dy}{dx} = \frac{\eta(x, y)}{\xi(x, y)} = -\frac{2y}{x}. \tag{60}$$

The solution is $y = \frac{\text{constant}}{x^2}$. Thus $r = yx^2 = \text{constant}$, and $s(r, x) = \left(\int \frac{dx}{\varepsilon(x, y(r, x))}\right)_{r=r(x,y)} = \left(\int \frac{dx}{x}\right)_{r=r(x,y)} = \ln|x|$. Thus our canonical coordinates are

$$(r, s) = (yx^2, \ln|x|). \tag{61}$$

Of course $x = e^s$ and $y = e^{-2s}r$. So $r_x = 2xy$; $r_y = x^2$; $s_x = \frac{1}{x}$; $s_y = 0$. Plug into equation (52):

$$\frac{ds}{dr} = \frac{s_x + \omega(x, y)s_y}{r_x + \omega(x, y)r_y} = \frac{\frac{1}{x}}{2xy - \omega(x, y)x^2} = \frac{\frac{1}{x}}{2xy - x^3y^2 - 2xy - \frac{1}{x}} = \frac{1}{x^4y^2 - 1} = \frac{1}{r^2 - 1}. \tag{62}$$

The Riccati equation has been reduced to quadrature in the canonical coordinates. That is

$$s(r) = \left(\int \frac{dr}{r^2 - 1}\right)_{r=x^2y} = \frac{1}{2}\ln\frac{r - 1}{r + 1}. \tag{63}$$

Converting back to the original coordinates we get

$$\ln|x| - \frac{1}{2}\ln\left|\frac{x^2y - 1}{x^2y + 1}\right| = \ln|x| + \ln\left|\sqrt{\frac{x^2y - 1}{x^2y + 1}}\right| = \ln\left|\sqrt{\frac{x^2y - 1}{x^2y + 1}} \cdot x\right| = c_0. \tag{64}$$

$$\frac{x^2y - 1}{x^2y + 1} \cdot x^2 = e^{c_0} = c. \tag{65}$$

$$x^4y - x^2 = cx^2y + c. \tag{66}$$

$$(x^4 - cx^2)y = c + x^2. \tag{67}$$

$$y = \frac{c + x^2}{x^2(x^2 - c)}. \tag{68}$$

We have solved the Riccati equation. We can get back the two solutions that were derived using the reduced characteristic equation by taking limits: $\lim_{c\to\infty} y = \frac{1}{x^2}$ and $\lim_{c\to 0} y = -\frac{1}{x^2}$. Note for sticklers, the "Riccati" equation is actually any ODE that is quadratic in the unknown function. It is nonlinear. We can solve it! Our method is general. Screw botany. Another note: looking at patterns that are invariant to symmetry $(\hat{x}, \hat{y}) = (e^{\epsilon}x, e^{-2\epsilon}y)$, I noticed that what we did would work for the Riccati equation with the following extra terms:

$$y' = Riccati\ equation + x^3y^3,$$

$$y' = Riccati\ equation + x^3y^3 + x^7y^5,$$

and so on. For the latter equation with the two extra terms we get, for example,

$$\frac{ds}{dr} = \frac{\frac{1}{x}}{2xy - \omega \cdot x^2} = \frac{\frac{1}{x^2}}{2y - x^2y^2 - 2y - \frac{1}{x^2} + x^4y^3 + x^8y^5} = \frac{\frac{1}{x^2}}{-x^2y^2 - \frac{1}{x^2} + x^4y^3 + x^8y^5}.$$

Since $x = e^s$ and $y = rx^{-2} = e^{-2s}r$, we get

$$\frac{ds}{dr} = \frac{\frac{1}{x^2}}{-x^2r^2x^{-4} - \frac{1}{x^2} + x^4r^3x^{-6} + x^8r^5x^{10}} = \frac{e^{-2s}}{-e^{-2s}r^2 - e^{-2s} + e^{-2s}r^3 + e^{-2s}r^5}$$

$$= \frac{1}{-(r^2 + 1 - r^3 - r^5)} = \frac{1}{r^5 + r^3 - r^2 - 1}.$$

Linearized symmetry condition. So here is what we have so far. One method to find symmetries of $\frac{dy}{dx} = \omega(x, y)$ is to use the symmetry condition (constraint)

$$\frac{\hat{y}_x + \omega(x,y)\hat{y}_y}{\hat{x}_x + \omega(x,y)\hat{x}_y} = \omega(\hat{x},\hat{y}), \tag{18}$$

which is usually a complicated PDE in both $\hat{x}$ and $\hat{y}$. By definition, Lie symmetries are of the form

$$\hat{x} = x + \varepsilon\xi(x,y) + O(\varepsilon^2), \tag{32}$$

$$\hat{y} = y + \varepsilon\eta(x,y) + O(\varepsilon^2), \tag{33}$$

where $\xi(x,y)$ and $\eta(x,y)$ are smooth. Note that to first order in ε, $\hat{y}_x = 0 + \varepsilon\eta_x$, $\hat{y}_y = 1 + \varepsilon\eta_y$, $\hat{x}_x = 1 + \varepsilon\varepsilon_x$, $\hat{x}_y = 0 + \varepsilon\xi_{y.}$ So when we substitute (32) and (33) into LHS of (18) we get:

$$\frac{\frac{\partial}{\partial x}\big(y + \varepsilon\eta(x,y)\big) + \omega(x,y)\frac{\partial}{\partial y}\big(y + \varepsilon\eta(x,y)\big)}{\frac{\partial}{\partial x}\big(x + \varepsilon\xi(x,y)\big) + \omega(x,y)\frac{\partial}{\partial y}\big(x + \varepsilon\xi(x,y)\big)} \tag{68}$$

$$= \frac{\varepsilon\eta_x + \omega(x,y)\big(1 + \varepsilon\eta_y\big)}{1 + \varepsilon\xi_x + \varepsilon\omega(x,y)\xi_y}.$$

Recall that $\frac{1}{1+\varepsilon x} \approx 1 - \varepsilon x$ when εx is small. Applying this binomial approximation to (68), we get

$$(68) = \big(\varepsilon\eta_x + \omega(x,y) + \omega(x,y)\varepsilon\eta_y\big)\big(1 - \varepsilon\xi_x + \varepsilon\omega(x,y)\xi_y\big)$$

$$= \varepsilon\eta_x + \omega(x,y) + \omega(x,y)\varepsilon\eta_y - \varepsilon^2\eta_x\xi_x - \varepsilon\xi_x\omega(x,y) + \varepsilon^2\eta_y\xi_y\omega^2(x,y) + \varepsilon^2\eta_x\xi_y\omega(x,y)$$

$$+ \varepsilon\omega^2(x,y)\xi_y + \varepsilon^2\eta_y\xi_y\omega^2(x,y),$$

and dropping terms higher than first order in ε as negligible we get

$$= \varepsilon\eta_x + \omega(x,y) + \omega(x,y)\varepsilon\eta_y - \varepsilon\xi_x\omega(x,y) + \varepsilon\omega^2(x,y)\xi_y.$$

Substituting (32) and (33) into the RHS)of (18) we get:

$$\omega(\hat{x},\hat{y}) = \omega(x + \varepsilon\xi, y + \varepsilon\eta) = \omega(x,y) + \varepsilon\xi\omega_x(x,y) + \varepsilon\eta\omega_y(x,y).$$

Putting the LHS together with the RHS we get:

$$\varepsilon\eta_x + \omega(x,y) + \omega(x,y)\varepsilon\eta_y - \varepsilon\xi_x\omega(x,y) + \varepsilon\omega^2(x,y)\xi_y = \omega(x,y) + \varepsilon\xi\omega_x(x,y) + \varepsilon\eta\omega_y(x,y).$$

Canceling things out and getting rid of ε, we get:

$$\eta_x + \omega(x,y)\eta_y - \xi_x\omega(x,y) + \omega^2(x,y)\xi_y = \xi\omega_x(x,y) + \eta\omega_y(x,y). \tag{69}$$

Finally rearranging, we get the ***linearized symmetry condition***

$$\eta_x + \left(\eta_y - \xi_x\right)\omega(x,y) - \omega^2(x,y)\xi_y = \xi\omega_x(x,y) + \eta\omega_y(x,y). \tag{70}$$

The linearized symmetry condition is a single PDE in two independent variables with infinitely many solutions, but it is linear and simpler that the original, nonlinearized PDE.

<u>Example 1.9</u>—Let's do it! Consider,

$$\frac{dy}{dx} = \frac{1-y^2}{xy} + 1. \tag{71}$$

From experience with these symmetry techniques, beginning with simpler differential equations and progressing onwards (much like Feynman did with his Feynman diagrams), our ansatz shall be:

$$\xi = \alpha(x), \quad \eta = \beta(x)y + \gamma(x). \tag{72}$$

We plug our ansatz into the linearized symmetry condition to get

$$\beta' y + \gamma'(x) + (\beta - \alpha')\left(\frac{1-y^2}{xy} + 1\right) = \alpha\left(\frac{y^2-1}{x^2y}\right) - (\beta y + \gamma)\left(\frac{1+y^2}{xy^2}\right). \tag{73}$$

Let's split (73) into a system of over determined equations by matching powers of y. On the LHS of (73) there are no terms with y^{-2}. On the RHS there is a term γ/xy^2. Then $\gamma = 0$. So (73) reduces to:

$$\beta' y + (\beta - \alpha')\left(\frac{1-y^2}{xy} + 1\right) = \alpha\left(\frac{y^2-1}{x^2y}\right) - \beta y\left(\frac{1+y^2}{xy^2}\right). \tag{74}$$

Matching LHS terms with y^{-2} to RHS with y^{-2} leads to

$$\frac{(\beta - \alpha')}{x} = \frac{\alpha}{x^2} - \frac{\beta}{x}. \tag{75}$$

Finally, matching LHS terms to RHS terms with y^0 leads to $(\beta - \alpha') = 0$, so $\beta = \alpha'$. Then the LHS of (75) equals zero. Equation (75) reduces to

$$0 = \frac{\alpha}{x^2} - \frac{\alpha'}{x}. \tag{76}$$

So $\alpha' + \frac{\alpha}{x} = 0$. Solving the simple ODE leads to $\alpha = c_1 x^{-1}$. This in turn tells us $\beta = c_1 x^{-2}$. Thus, finally, we have $\xi(x) = c_1 x^{-1}$, and $\eta(x, y) = c_1 x^{-2} y$. We have our tangent vector. So far to me it does not appear that this symmetry came from a translation symmetry, so I have not found canonical coordinates. However the reduced characteristic does lead to solutions. Recall that $\tilde{Q}(x, y, \omega(x, y)) = \eta(x, y) - \omega(x, y)\xi(x, y)$. Substituting $\omega(x, y), \xi(x)$, and $\eta(x, y)$, we get

$$\tilde{Q}(x, y, \omega(x, y)) = -c_1 \frac{y}{x^2} - \left(\frac{1 - y^2}{xy} + 1\right)\frac{c_1}{x} \tag{77}$$

$$= -\frac{{c_1}^2}{x^2}\left(y + \frac{1}{y} \cdot (1 - y^2 + xy)\right).$$

If we set $\tilde{Q}$ to zero we get:

$$y + \frac{1}{y} \cdot (1 - y^2 + xy) = 0. \tag{78}$$

This is so if $y = -\frac{1}{x}$. We check this by substituting the solution into both the RHS and LHS of (71) to get $\frac{dy}{dx} = \frac{1-y^2}{xy} + 1 = \frac{1-\frac{1}{x^2}}{-1} + 1 = \frac{1}{x^2} = \frac{d}{dx}y = -\frac{d}{dx}\frac{1}{x}$.

Let's write the reduced characteristic in terms of the linearized symmetry condition as follows:

$$\tilde{Q} = \eta - \omega\xi. \tag{79}$$

Let

$$\tilde{Q}_x + \omega\tilde{Q}_y = \omega_y \tilde{Q}. \tag{80}$$

Now let's take the appropriate partial derivatives of (79).

$$\eta_x - \omega_x\xi - \omega\xi_x + \omega\left(\eta_y - \omega_y\xi - \omega\xi_y\right) = \omega_y(\eta - \omega\xi). \tag{81}$$

$$\eta_x - \omega_x\xi - \omega\xi_x + \omega\eta_y - \omega\omega_y\xi - \omega^2\xi_y = \omega_y\eta - \omega_y\omega\xi. \tag{82}$$

$$\eta_x - \omega_x\xi - \omega\xi_x + \omega\eta_y - \omega^2\xi_y = \omega_y\eta. \tag{83}$$

Rearranging leads to the linearized symmetry condition:

$$\eta_x + \left(\eta_y - \xi_x\right)\omega - \xi_y\omega^2 = \xi\omega_x + \eta\omega_y. \tag{70}$$

If $\tilde{Q}$ satisfies (79), then $(\xi, \eta) = \left(\xi, \tilde{Q} + \omega\xi\right)$ is a tangent vector field of a one-parameter group. All Lie symmetries correspond to the solution $\tilde{Q} = 0$. Nontrivial symmetries can be found from (80) using the method of characteristics

$$\frac{dx}{1} = \frac{dy}{\omega(x,y)} = \frac{d\tilde{Q}}{\omega(x,y)\tilde{Q}}.$$

The LHS is, uselessly, our original ODE. Lastly note that if (ξ, η) is a nonzero solution to the linearized symmetry condition, then so is $(k\xi, k\eta)$ $\forall\, k > 0$. This freedom corresponds to replacing ε by $k^{-1}\varepsilon$, which does not alter the orbits of the Lie group. So the same Lie symmetries are recovered, irrespective of the value of k. The freedom to rescale ε allows us to multiply $\tilde{Q}$ by any nonzero constant without altering the orbits.

On patterns—We may take derivatives by always applying the formal definition of a derivative, $e.g.$, $\frac{d}{dx}x^2 = \lim_{\Delta X \to 0} \frac{(X+\Delta X)^2 - X^2}{\Delta X} = 2x$, but if we take enough derivatives we begin to discover patterns such as the power rule, the quotient rule, or the chain rule. The same applies to using symmetry methods to extract solutions to differential equations. Some common symmetries, including translations, scalings and rotations can be found with the ansatz

$$\xi = c_1x + c_xy + c_{3,}\ \eta = c_4x + c_5y + c_6 \tag{84}$$

This ansatz is more restrictive than ansatz (72). Ansatz (84) works for

$$y' = \frac{y - 4xy - 16x^2}{y^3 + 4x^2y + x}$$

if $c_1 = c_3 = c_5 = c_6 = 0,\ c_4 = -4c_2$. Specialized computer algebra packages have been created to assist with symmetry methods for differential equations. For first order ODEs the search for a nontrivial symmetry may be fruitless even though the ODE might have infinitely many symmetries. Symmetries of higher order ODEs and PDEs can usually be found systematically. The following link from MapleSoft is a tool for finding symmetries for differential equations (July 2012:

http://www.maplesoft.com/support/help/Maple/view.aspx?path=DEtools/symgen

Please refer to chapter two of the Hydon textbook for many comparisons and relationships between symmetry methods and standard methods.

Infinitesimal generator. Suppose a first order ODE has a one-parameter Lie group of symmetries whose tangent vector at (x, y) is (ε, η). Then the partial differential operator

$$X = \xi(x, y)\partial_x + \eta(x, y)\partial_y \tag{85}$$

Is the *infinitesimal generator* of the Lie group. We have already encountered and used such infinitesimal generators. Recall

$$\xi(x, y)r_x + \eta(x, y)r_y = 0, \tag{40}$$

$$\xi(x, y)s_x + \eta(x, y)s_y = 1. \tag{41}$$

We may rewrite them as

$$X_r = 0, \tag{86}$$

$$X_s = 1. \tag{87}$$

We shall soon see that the algebraic properties of the infinitesimal generators of a differential equation (under commutation) will tell us if we can reduce the order of the differential equation by one or more.

In the following examples let's suppose that we have a symmetry of some differential equation. From this symmetry we compute the corresponding infinitesimal generator.

Example 1.10—For the Riccati equation, one symmetry is $(\hat{x},\hat{y})=(e^{\varepsilon}x,e^{-2\varepsilon}y)$. Then, $\xi(x,y)=\left(\frac{d\hat{x}}{d\varepsilon}\right)_{\varepsilon=0}=(e^{\varepsilon}x)_{\varepsilon=0}=x$ and $\eta(x,y)=\left(\frac{d\hat{y}}{d\varepsilon}\right)_{\varepsilon=0}=(-2e^{-2\varepsilon}y)_{\varepsilon=0}=-2y$. Therefore $X=x\partial_x-2y\partial_y$. The canonical coordinates were $(r,s)=(yx^2,\ln|x|)$.

Example 1.11—The differential equation $\frac{dy}{dx}=\frac{y+1}{x}+\frac{y^2}{x^3}$ has Lie symmetries of the form $(\hat{x},\hat{y})=\left(\frac{x}{1-\varepsilon x},\frac{y}{1-\varepsilon x}\right)$. The tangent vector is given by (x^2,xy). Therefore $X=x^2\partial_x+xy\partial_y$.

Example 1.12—Given $(\hat{x},\hat{y})=(x+\varepsilon,y+\varepsilon)$, $\xi(x,y)=\left(\frac{d\hat{x}}{d\varepsilon}\right)_{\varepsilon=0}=1$ and $\eta(x,y)=\left(\frac{d\hat{y}}{d\varepsilon}\right)_{\varepsilon=0}=1$. Therefore $X=\partial_x+\partial_y$.

Optional—6 Page Double Spaced "BIG PICTURE" Motivational digression Example with Discussion

Example 1.13—Let's peer ahead into the extreme importance of understanding the key algebraic properties of the infinitesimal generators of a differential equation to mathematicians and physicists. Don't worry. We will build soon build the math in detail. Consider the following fourth order ODE

$$y^{(iv)}=y'''^{\frac{4}{3}}. \tag{88}$$

By applying the linearized symmetry condition for n^{th} order ODEs (see chapter 2) we get the following set of infinitesimal generators associated with ODE (88): $X_1=\partial y,\ X_2=x\partial y,\ X_3=x^2\partial y,\ X_4=\partial x$, $X_5=x\partial x$. The ***commutators*** of these infinitesimal generators are defined by:

$$[X_i,X_j]=X_iX_j-X_jX_i. \tag{89}$$

If $i=j$ the commutator is zero. Some other commutators with $i\neq j$ are zero. Some are not zero, *e.g.,*

$$[X_1, X_2] = X_1X_2 - X_2X_1 = \partial y(x\partial y) - x\partial y(\partial y) = x - x = 0, \quad (90)$$

$$[X_2, X_4] = X_2X_4 - X_4X_2 = x\partial y(\partial x) - \partial x(x\partial y) = 0 - \partial y = -X_1. \quad (91)$$

If you try all the remaining possible pairs of commutators, the only other nonzero commutators are

$$[X_2, X_5] = -X_2, \qquad [X_3, X_4] = -2X_2, \qquad [X_3, X_5] = -2X_3, \qquad [X_4, X_5] = X_4.$$

ODE (88) has a five-dimensional ***Lie algebra*** $\mathcal{L} = \{X_1, X_2, X_3, X_4, X_5\}$. Note that from all possible commutators taken from the set $\{X_1, X_2, X_3, X_4, X_5\}$ we only get back the infinitesimal generators $\{X_1, X_2, X_3, X_4\}$. There is no $[X_i, X_j] = \mathrm{c} \cdot X_5$ where c is nonzero. We say that the infinitesimal generators $\{X_1, X_2, X_3, X_4, X_5\}$ form a ***derived subalgebra*** $\mathcal{L}^{(1)} = \mathrm{span}(X_1, X_2, X_3, X_4)$ under the (binary) operation of commutation. Now by taking all possible commutators from the set $\{X_1, X_2, X_3, X_4\}$ you only get back two infinitesimal generators, namely, $\mathcal{L}^{(2)} = \mathrm{span}\{X_1, X_2\}$. Repeating this using all possible commutators from $\{X_1, X_2\}$, you get $\mathcal{L}^{(3)} = \{0\}$.

Let's sum up. Beginning with $\{X_1, X_2, X_3, X_4, X_5\}$, we built a "solvable tower" down to $\mathrm{span}\{X_1, X_2, X_3, X_4\}$, down to $\mathrm{span}\{X_1, X_2\}$, and finally down to $\mathrm{span}\{0\}$ under commutation. When this is possible, which isn't always, the subalgebra is said to be ***solvable***. This is important. An n^{th} order ODE with $R \leq n$ Lie point symmetries can be reduced in order by $n - R$, becoming an algebraic equation if $n = R$. If the ODE is of order R with an R-dimensional solvable Lie algebra, we can integrate the ODE in terms of differential invariants stepwise R times After taking tons of undergraduate and graduate algebra I knew of no good reason as to why I should care about solvability. My experience was not unique. Most PhD mathematicians have never learned the useful nature of solvability for differential equations, including the algebraists among them. Whatever it is they do know, it is incomplete and lacking defining context. A terrible consequence of this is that PhD physicists, who go through school eating and breathing differential equations only learn to "ape" seemingly disparate algebraic methods, coming out ignorant of very powerful and unifying symmetry methods, necessarily suffering the consequences of having a confused, disconnected collection of magical mathematical prescriptions, namely suffering a needless weakening of their fundamental understanding of physics. Most of us are not keen to waste time with useless mental masturbation and dissipation. To the greatest extent possible, people do not go into physics to practice witchcraft in terms of magical mathematical prescriptions. Let mathematics be well taught. Let mystery come from nature. By the end of Part I you will operate with the algebraic properties of infinitesimal generators and their commutation relationships to treat differential equations the way you might sum sports figures.

Let's now direct this digression to the physical importance of infinitesimal generators and their commutation algebra. When you study quantum physics you learn rules for converting classical physics equations into their quantum mechanical counterparts. We assign classical physics energy to a differential operator: $E \longmapsto i\hbar\frac{\partial}{\partial t}$. Sticking to just one space dimension, we assign the classical physics momentum to a differential operator: $p_x \longmapsto -i\hbar\frac{\partial}{\partial x}$. Position x and time t remain unaffected when they are transformed to operators: $x \longmapsto x,\ t \longmapsto t$. Quantum mechanical operators operate on wave functions $\psi(x,t)$. Let's show this with a simple, one-dimensional, classical physics spring-mass system. The total energy of a mass m oscillating at the end of a spring with spring constant k is $E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{p_x^2}{2m} + \frac{1}{2}kx^2$. Converting this equation into a quantum mechanical equation leads to the Schrödinger equation for the quantum mechanical oscillator $i\hbar\frac{\partial}{\partial t}\,\psi(x,t) = \left[\frac{1}{2m}\left(i\hbar\frac{\partial}{\partial x}\right)^2 + \frac{1}{2}kx^2\right]\psi(x,t) = \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x,t) + \frac{1}{2}kx^2\psi(x,t)$, a crude model of an electron vibrating to and from a molecular or atomic nucleus giving rise to vibrational spectra. The commutator algebra of the differential operators we are using tells us what pairs of physical observables we can measure simultaneously with infinite precision, at least in principle, and what pairs of physical observables we cannot measure simultaneously with infinite precision no matter what we do.

Consider the commutator of the momentum and position operators, $\left[-i\hbar\frac{\partial}{\partial x}, x\right]\psi(x,t)$. Using the chain rule, we have

$$\left[-i\hbar\frac{\partial}{\partial x}, x\right]\psi(x,t) = -i\hbar\frac{\partial}{\partial x}\big(x\psi(x,t)\big) - x\left(-i\hbar\frac{\partial}{\partial x}\right)\psi(x,t)$$

$$= -i\hbar\psi(x,t) - i\hbar x\frac{\partial\psi(x,t)}{\partial x} + i\hbar x\frac{\partial\psi(x,t)}{\partial x} = -i\hbar\psi(x,t).$$

Abusing notation, we write $\left[-i\hbar\frac{\partial}{\partial x}, x\right] = -i\hbar$. This commutator is not zero. Position and momentum do not commute. As a consequence of this, the uncertainty principle applies to position and momentum. The uncertainty in position Δx and the uncertainty in momentum Δp obey $\Delta x\Delta p \geq \frac{\hbar}{2}$. The more precisely we measure one of these observables, the more uncertainty shrouds the other observable. On the other hand, if the commutator had been zero, we could simultaneously measure position and momentum with infinite precession, at least in principle. This paragraph was never meant to replace a course in quantum physics, but to demonstrate the potential physical meaning of the derived commutator algebra of a physics-based differential equation.

Sometimes the algebraic structure of the infinitesimal generators of a differential equation result in a discrete group under commutation. Such discrete groups can always be mimicked by matrices. In linear algebra you study, among other important things, how to compute the eigenvalues and eigenvectors of matrices. In physics, these eigenvalues can correspond to the particle spectra of the Standard Model of our universe, or to the particle spectra of some theoretical extension of the Standard Model, or even to the particle spectra of some hypothetical universe. **In a nutshell, the process of studying (theoretical) universes begins with deriving the differential equations of motion (via the calculus of variations in terms of an "action principle" (Part II)). If the infinitesimal generators of a differential equation result in a discrete group either directly or from a higher subalgebra, we learn**

how far the differential equation may be reduced in order; we learn which observables are subject to the uncertainty principle; there is, additionally, an associated matrix group that mimics the algebraic structure of the commutator relationships of the discrete group of infinitesimal generators of the differential equations. This group-theoretic matrix structure yields eigenvalues and eigenvectors that may correspond to the particle spectra of some universe (Part IV).

Strangely, not everything is particle physics. If we're talking about electrons orbiting nuclei, eigenstates and eigenvectors are their energies and orbitals. Clearly what the eigenstates represent depends on what we are modeling. The Gell-Mann matrices (which mimic the discrete group structure of the commutator relationships of the infinitesimal generators of a differential equation derived from a variational action principle associated with a model of the strong nuclear force) are:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

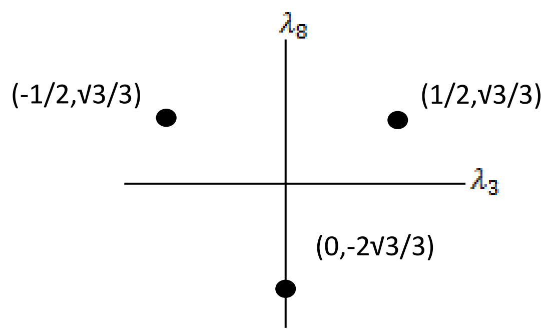
$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

These matrices are hermitian. That is they equal their complex conjugate transpose $\lambda_i = \lambda_i^\dagger$ and they are traceless (the sum of their diagonal elements equals 0). As usual, we define the commutators as $[\lambda_a, \lambda_b] = \lambda_a\lambda_b - \lambda_b\lambda_a$. Here are a few nonzero cases: $[\lambda_1, \lambda_2] = 2i\lambda_{3,}$ $[\lambda_2, \lambda_1] = -2i\lambda_3$. In general, $\left[\lambda_i, \lambda_j\right] = \frac{1}{2} i f_{ijk}, \lambda_k,$ where the repeated index implies summation, $\left[\lambda_i, \lambda_j\right] = \frac{1}{2} i\big(f_{ij1}, \lambda_1 + f_{ij2}, \lambda_{2+\dots} + f_{ij8}, \lambda_8\big)$. The only nonzero f_{ijk} are $f_{123} = 1,$ $f_{458} = f_{678}\frac{\sqrt{3}}{2},$ $f_{147} = f_{165} = f_{257} = f_{345} = f_{376} = \frac{1}{2}.$ Any f_{ijk} involving a permutation of i, j, and k not on the list is zero. Even permutations, *e.g.*, 123→312→213 have the same value. The odd permutations, *e.g.*, 123→132 have opposite sign. That is, $f_{123} = f_{312} = f_{231} = 1$, while $f_{123} = f_{321} = f_{213} = -1$. The symbol f_{ijk} is said to be totally antisymmetric in i, j, and k. For example, $[\lambda_4, \lambda_8] = 2if_{45a}\lambda_a = 2if_{451}\lambda_1 + \cdots + 2if_{453}\lambda_4 + \cdots + 2if_{458}\lambda_8 = 0 + 0 + 2i \cdot \frac{1}{2} f_{453}\lambda_3 + 0 + 0 + 0 + 0 + 2i \cdot \frac{\sqrt{3}}{2} f_{458}\lambda_8 = 2i \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$. You may verify this by checking $\lambda_4\lambda_8 - \lambda_8\lambda_4$ directly. Note that λ_3 and λ_8 are diagonal. Hence they commute. That is, $\lambda_3\lambda_8 - \lambda_8\lambda_3 = 0$. Quantum physics tells us that these therefore have simultaneously measurable eigenvalues. The simultaneous eigenvectors of these two matrices are

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

with eigenvalues $\{1, -1, 0\}$ for λ_3 and $\left\{\frac{\sqrt{3}}{3}, -\frac{\sqrt{3}}{3}, -\frac{2\sqrt{3}}{3}\right\}$ for λ_8. If we plot these against each other



we obtain a diagram very much resembling how the up, down, and strange quarks show up on a graph of hypercharge.

At this point in these notes, this paragraph is for advanced graduate students in physics. By the end of these notes, however, this paragraph should make much more sense to all readers. Given the above, doesn't it make sense to first learn about how to use action principles and the calculus of variations together with Noether's theorem to derive the differential equations of physics, then study these differential equations using symmetry methods to find any Lie symmetries and the corresponding discrete commutator algebra of their infinitesimal generator realizations if any? The derived subalgebra may go down to $\mathrm{span}\{0\}$ under commutation as in example 1.13, or it may stop shy of $\mathrm{span}\{0\}$, forming a discrete group (such as SU(2)) which can be copied by matrix representations with the same group structure. We learn what is simultaneously measureable in terms of the eigenvalues and eigenvectors of the representation matrices. Note that while the group may be discrete, the infinitesimal generators themselves are continuous, the underlying Lie point symmetries being continuous groups. If the X_i form a linear vector space, then the local, continuous properties of these infinitesimal generators under commutation, *e.g.*, $[X_2, X_4] = 0 \text{ or } \neq 0$, relate to the measure of the curvature of a space in general relativity (in terms of the group of all coordinate transformations and connection coefficients), and similarly to the gauge filed vector potentials in quantum field theories (in both cases in terms of the commutators of the respective covariant derivatives). This unified overview should not take a decade to master beyond earning a doctoral degree in physics, as it did with me. It's physical underpinnings and mathematical grammar should be learned by the undergraduate between the sophomore and senior year, to be fully mastered during the first year and a half of graduate school where one should be learning the details of advanced mechanics, electrodynamics, quantum physics, thermodynamics and statistical physics, quantum electrodynamics, relativity, and possibly quantum fields with the power of a unified mathematical grammar and physical overview of where shit fits together and why it fits together. (Note—It doesn't generally go backwards uniquely from postulating matrix representations back to (infinitesimal generator) realizations, back to differential equations, but this is nevertheless also a valid approach to investigate theoretical universes and extensions beyond the Standard Model.)

There is, of course, more mathematics and physics beyond the unified outline expounded on in this work, but it ties back. For now, let's get back to the ground. We still have to learn to crawl. After finishing up symmetry methods for first order ODEs in chapter 1, we shall proceed to extend (or “prolong”) these symmetry methods to higher order linear and nonlinear ODEs, and finally to linear and nonlinear PDEs.

END Optional—6 Page Double Spaced “BIG PICTURE” Motivational digression Example with Discussion

Change of coordinates and the infinitesimal generator. How is the infinitesimal generator affected by a change of coordinates? Suppose (u, v) are new coordinates and let $F(u, v)$ be an arbitrary smooth function. By the chain rule

$$XF(u,v) = XF\big(u(x,y), v(x,y)\big) = \xi(x,y)\partial_x F + \eta(x,y)\partial_y F$$

$$= \xi\left[\frac{\partial F}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial x}\right] + \eta\left[\frac{\partial F}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial y}\right] = \xi[u_x F_u + v_x F_v] + \eta\big[u_y F_u + v_y F_v\big]$$

$$= \big(\xi u_x + \eta u_y\big)F_u + \big(\xi v_x + \eta v_y\big)F_v = (Xu)F_u + (Xv)F_v.$$

Without loss of generality, since $F(u, v)$ is arbitrary, in the new coordinates

$$X = (Xu)\,\partial_u + (Xv)\,\partial_{v.}$$

Thus X represents the tangent vector field in all coordinate systems. If we regard $\{\partial_x, \partial_y\}$ as a *basis* for the space of vector fields on the plane, X is the tangent vector at (x, y). The infinitesimal generator provides a coordinate free way of characterizing the action of Lie symmetries on functions.

If $(u, v) = (r, s)$ are canonical coordinates, the tangent vector is $(0,1)$ and $X = \partial_s$. Let $G\big(r(x,y), s(x,y)\big)$ be a smooth function and $F(x,y) = G\big(r(x,y), s(x,y)\big)$. At any invariant point (x, y), the Lie symmetries map $F(x, y)$ to $F(\hat{x},\hat{y}) = G(\hat{r},\hat{s}) = G(r, s+\varepsilon)$. Applying Taylor’s theorem and given $X = \partial_s$, we get

$$F(\hat{x},\hat{y}) = \sum_{j=0}^{\infty}\frac{\varepsilon^j}{j!}\frac{\partial^j G(r,s)}{\partial s^j} = \sum_{j=0}^{\infty}\frac{\varepsilon^j}{j!}X^j G(r,s).$$

Reverting back to (x, y) coordinates, $F(\hat{x}, \hat{y}) = \sum_{j=0}^{\infty} \frac{\varepsilon^j}{j!} X^j F(x, y)$. If the series converges it is called the Lie series of F about (s, y). We have assumed that (s, y) is not an invariant point, but the expansion is also valid at all invariant points. At an invariant point $X = 0$, and only the $j = 0$ term survives, which is $F(x, y)$. We may express all of this in shorthand to $F(\hat{x}, \hat{y}) = F(e^{\varepsilon X} x, e^{\varepsilon X} y) = e^{\varepsilon X} F(x, y)$.

<u>Example 1.14</u>—Ever wondered what $e^{\theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}}$ means? Recall $e^x = 1 + x + \frac{1}{2!} x^2 + \frac{1}{3!} x^3 + \cdots$, but now x is a matrix. The expression therefore means

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \theta + \frac{1}{2!} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 \theta^2 + \frac{1}{3!} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^3 \theta^3 + \cdots$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \theta + \frac{1}{2} \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \theta^2 + \frac{1}{6} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \theta^3 + \cdots$$

$$= \begin{pmatrix} 1 - \frac{1}{2}\theta^2 + \cdots & -\theta + \cdots \\ \theta + \cdots & 1 - \frac{1}{2}\theta^2 + \cdots \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = R(\theta).$$

So $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is the infinitesimal generator of rotations of the *xy*-plane. That is $\left(\frac{dR(\theta)}{d\theta}\right)_{\theta=0} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. A finite rotation ϑ is generated by the Taylor series expansion of $e^{\theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}}$. $R(\theta)$is a continuous group in θ.

Lie symmetries can be reconstructed as $\hat{x} = e^{\varepsilon X} x,\ \hat{y} = e^{\varepsilon X} y$. Thus $F(e^{\varepsilon X} x, e^{\varepsilon X} y) = e^{\varepsilon X} F(x, y)$. This generalizes to *L* variables, $z^1, \ldots, z^L$ where the Lie symmetries are $\hat{z}^s(z^1, \ldots, z^L; \varepsilon) = z^s + \varepsilon(z^1, \ldots, z^L) + O(\varepsilon^2),\ s = 1, \ldots, L$. Then the one-parameter Lie group is $X = \zeta^s(z^1, \ldots, z^L) \frac{\partial}{\partial z^s}$. Note that (Einstein) summation convention is used. If an index is used twice, then sum over all possible values of that index. Lie symmetries may be reconstructed from the Lie series $\hat{z}^s = e^{\varepsilon X} z^s$. If F is a smooth function, $F(e^{\varepsilon X} z^1, \ldots, e^{\varepsilon X} z^L) = e^{\varepsilon X} F \zeta^s(z^1, \ldots, z^L)$.

Note that it is only sometimes easy to work from an infinitesimal generator back to a one-parameter Lie group. Consider for example $X = \partial_x + y\partial_y$. Then $\xi(x, y) = 1$ and $\eta(x, y) = y$. Looking at the definitions, $\left(\frac{d\hat{x}}{d\varepsilon}\right)_{\varepsilon=0} = 1$ and $\left(\frac{d\hat{y}}{d\varepsilon}\right)_{\varepsilon=0} = y$. Thus by visual inspection $\hat{x} = x + \varepsilon$ and $\hat{y} = e^{\varepsilon}y$ work, and this is our Lie symmetry for our infinitesimal generator.

A parting example 1.15 (optional)—In classical mechanics angular momentum is defined by

$$L = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ x & y & z \\ p_x & p_y & p_z \end{vmatrix} = \hat{\imath}(yp_z - zp_y) - \hat{\jmath}(xp_y - yp_x) + \hat{k}(xp_y - yp_x) = L_x + L_y + L_z.$$

In quantum physics we turn these quantities into differential operators. Recall from the digression that $p_x \longmapsto -i\hbar\frac{\partial}{\partial x}$. Then, for example, $L_x \to \hat{L}_x = -i\hbar\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right)$, $L_y \to \hat{L}_y = -i\hbar\left(x\frac{\partial}{\partial z} - z\frac{\partial}{\partial x}\right)$, and $L_z \to \hat{L}_z = -i\hbar\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)$. Let's study the commutators of $\{\hat{L}_x, \hat{L}_y, \hat{L}_z\}$, e.g.,

$$[\hat{L}_x, \hat{L}_y] = \left[-i\hbar\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right)\cdot -i\hbar\left(x\frac{\partial}{\partial z} - z\frac{\partial}{\partial x}\right) - -i\hbar\left(x\frac{\partial}{\partial z} - z\frac{\partial}{\partial x}\right)\cdot -i\hbar\cdot\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right)\right]$$

$$= -\hbar^2\left[\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right)\cdot\left(x\frac{\partial}{\partial z} - z\frac{\partial}{\partial x}\right) - \left(x\frac{\partial}{\partial z} - z\frac{\partial}{\partial x}\right)\cdot\left(y\frac{\partial}{\partial z} - z\frac{\partial}{\partial y}\right)\right]$$

$$= -\hbar^2\left[y\frac{\partial}{\partial z}\left(x\frac{\partial}{\partial z}\right) - y\frac{\partial}{\partial z}\left(z\frac{\partial}{\partial x}\right) - z\frac{\partial}{\partial y}\left(x\frac{\partial}{\partial z}\right) + z\frac{\partial}{\partial y}\left(z\frac{\partial}{\partial x}\right) - x\frac{\partial}{\partial z}\left(y\frac{\partial}{\partial z}\right)\right.$$

$$\left.+ x\frac{\partial}{\partial z}\left(z\frac{\partial}{\partial y}\right) + z\frac{\partial}{\partial x}\left(y\frac{\partial}{\partial z}\right) - z\frac{\partial}{\partial x}\left(z\frac{\partial}{\partial y}\right)\right]$$

$$= -\hbar^2\left[xy\frac{\partial^2}{\partial z^2} - y\frac{\partial}{\partial x} - yz\frac{\partial^2}{\partial x\partial z} - xz\frac{\partial^2}{\partial y\partial z} + z^2\frac{\partial^2}{\partial x\partial y} - xy\frac{\partial^2}{\partial z^2} + x\frac{\partial}{\partial y} + xz\frac{\partial^2}{\partial y\partial z}\right.$$

$$\left.+ yz\frac{\partial^2}{\partial x\partial z} - z^2\frac{\partial^2}{\partial x\partial y}\right] = -\hbar^2\left[-y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}\right] = -\hbar^2 L_z.$$

There is an easier way: the rotation matrices have the same commutator algebra.

$$R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix},\ R_y(\psi) = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix},\ R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Taylor series expand the matrix elements to first order to get the infinitesimal generator versions:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The factor i is put in to match the signs with the angular momentum commutator algebra up to a constant, *e.g.*, $[R_x, R_y] = -R_z$. The algebra of commutators formed by the rotation matrices under commutation defines a discrete group, $SO(3)$. The matrices themselves have continuous parameters corresponding to continuous groups.

When working with differential operators we speak of ***realizations*** of a symmetry. When working with matrices, we speak of ***representations*** of a symmetry. As we shall see later, it is often better to work with representations than realizations. Matrix representations are not necessarily unique though. The 2×2 Pauli matrices have the same commutator algebra up to a complex constant.

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

They represent the group $SU(2)$ (special unitary group). The eigenvectors of σ_z are associated with electron spin, as opposed to rotations about an origin in classical physics. We can map the elements of $SU(2)$ to $O(3)$ as well as the negatives of the elements of $SU(2)$ to $SO(3)$, a two-to-one mapping. When we get to algebraic topology and consider the geometric meaning of commutators, we will see that not having a 1-1 mapping between two groups tells us that they correspond to globally different spaces, much like a sphere and torus are topologically different spaces (manifolds). The geometric meaning I allude to unifies the picture of quantum fields and general relativity. Let's attack higher order ODEs!

Part I. **Chapter 2**. (Accessible to sophomores; required for mathematics/physics majors up through the postdoctoral research level) On to higher order ODEs. (First full example for 2nd order ODE at fifth page.)

We prolong the symmetry tools for first order ODEs to higher order ODEs. Consider

$$y^{(n)} = \omega\big(x, y, y', \dots, y^{(n-1)}\big), \qquad n \geq 2, \qquad y^{(k)} = \frac{d^k y}{dx^k}, \tag{1}$$

Where $\omega$is locally smooth in its arguments. A symmetry is a diffeomorphism mapping the set of ODE solutions to itself. Any diffeomorphism $\Gamma: (x, y) \longmapsto (\hat{x}, \hat{y})$maps smooth planar curves to smooth planar curves. This action of Γ on the plane induces an action on the derivatives $y^{(k)}$, $\Gamma: \big(x, y, y', \dots, y^{(n)}\big) \longmapsto \big(\hat{x}, \hat{y}, \hat{y}', \dots, \hat{y}^{(n)}\big)$, where $\hat{y}^{(k)} = \frac{d^k \hat{y}}{d\hat{x}^k}$, $k = 1, \dots, n$, called the n^{th} prolongation of Γ.

The functions $\hat{y}^{(k)}$ are calculated recursively using the chain rule as follows:

$$\hat{y}^{(k)} = \frac{d\hat{y}^{(k-1)}}{dx^{k-1}} = \frac{D_x \hat{y}^{(k-1)}}{D_x \hat{x}}, \qquad \hat{y}^0 = \hat{y}, \tag{2}$$

where $D_x = \partial_x + y'\partial_y + y''\partial_{y'} + \cdots$. The symmetry condition: $\hat{y}^{(n)} = \omega\big(\hat{x}, \hat{y}, \hat{y}', \dots, \hat{y}^{(n-1)}\big)$ when $y^{(n)} = \omega\big(x, y, y', \dots, y^{(n-1)}\big)$ is usually nonlinear. As with first order ODEs, Lie symmetries are obtained by linearization about $\varepsilon = 0$, which is not possible for discrete symmetries (such as reflections).

Example 2.1—$(\hat{x}, \hat{y}) = \left(\frac{1}{x}, \frac{y}{x}\right)$ is a symmetry of $y'' = 0$, $x > 0$. Let's check.

$$\hat{y}^{(1)} = \frac{d^0 \hat{y}}{d\hat{x}^0} = \frac{D_x\left(\frac{y}{x}\right)}{D_x\left(\frac{1}{x}\right)} = \frac{\left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\left(\frac{y}{x}\right)}{\left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\left(\frac{1}{x}\right)} = \frac{-\frac{y}{x^2} + y' \cdot \frac{1}{x}}{-\frac{1}{x^2}} = y - xy'.$$

$$\hat{y}^{(2)} = \frac{d^1 \hat{y}'}{d\hat{x}^1} = \frac{D_x(y - xy')}{D_x\left(\frac{1}{x}\right)} = \frac{\big(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\big)(y - xy')}{\big(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\big)\left(\frac{1}{x}\right)} = \frac{0 - y' + y' - y''x}{-\frac{1}{x^2}} = x^3 y''.$$

Therefore $\hat{y}'' = 0$ when $y'' = 0$. The symmetry is its own inverse thus it belongs to a group of order 2. The general solution, $y_{gen} = c_1 + c_x x$ gets mapped to $\hat{y} = \frac{y}{x} = c_1 + \frac{c_x}{x} = c_1 + c_2\hat{x}$. If you have had a little group theory you are probably familiar with the group containing $\{0,1\}$ with the binary operation being sum modulo 1, $\langle \mathbb{Z}_2, + \rangle$.

The linearized symmetry condition for Lie symmetries for higher order ODEs is derived by the same method used for first order ODEs. Given an ansatz symmetry $\Gamma: (x, y) \longmapsto (\hat{x}, \hat{y})$, the trivial symmetry corresponds to $\varepsilon = 0$. The prolonged Lie symmetries are of the form

$$\hat{x} = x + \varepsilon\xi + O(\varepsilon^2),$$

$$\hat{y} = y + \varepsilon\eta + O(\varepsilon^2), \tag{3}$$

$$\hat{y}^{(k)} = y^{(k)} + \varepsilon\eta^{(k)} + O(\varepsilon^2),$$

Where $k \geq 1$. The superscript $\eta^{(k)}$ is merely an index. Substitute (2) into $\hat{y}^{(n)} = \omega\left(\hat{x}, \hat{y}, \hat{y}', \dots, \hat{y}^{(n-1)}\right)$.

$$\omega\left(\hat{x}, \hat{y}, \hat{y}', \dots, \hat{y}^{(n-1)}\right) = \omega\left(x + \varepsilon\xi + O(\varepsilon^2), y + \varepsilon\eta + O(\varepsilon^2), y' + \varepsilon\eta^{(1)} + O(\varepsilon^2), \dots, y^{(n-1)} + \right.$$

$$\left. \varepsilon\eta^{(n-1)} + O(\varepsilon^2)\right) = \omega + \left[\xi\omega_x + \eta\omega_y\right]\varepsilon + \left[\eta^{(1)}\omega_{y'} + \cdots + \eta^{(n-1)}\omega_{y^{(n-\prime)}}\right]\varepsilon,$$

where the **linearized symmetry condition for an *n*th order ODE** is

$$\eta^{(n)} \equiv \xi\omega_x + \eta\omega_y + \eta^{(1)}\omega_{y'} + \cdots + \eta^{(n-1)}\omega_{y^{(n-\prime)}} \tag{4}$$

We find the $\eta^{(n)}$ recursively from (2). Recall $\frac{1}{1+x} \approx 1 - x$ for small x. Consider the case for $k = 1$:

$$\hat{y}^{(1)} = \frac{D_x\hat{y}^{(0)}}{D_x\hat{x}} = \frac{\left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\hat{y}}{\left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\hat{x}} = \frac{\left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\left(y + \varepsilon\eta + O(\varepsilon^2)\right)}{\left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\hat{x} = x + \varepsilon\xi + O(\varepsilon^2)}$$

$$= \frac{y' + \varepsilon D_x\eta + O(\varepsilon^2)}{1 + \varepsilon D_x\xi + O(\varepsilon^2)} = (y' + \varepsilon D_x\eta)(1 - \varepsilon D_x\xi) = y' + \varepsilon D_x\eta - y'\varepsilon D_x\xi + O(\varepsilon^2)$$

$$= y' + \varepsilon(D_x\eta - y'D_x\xi) + O(\varepsilon^2).$$

So,

$$\eta^{(1)} = (D_x\eta - y'D_x\xi) \tag{5}$$

Then continuing on to the k^{th} step, we have

$$\hat{y}^{(k)} = \frac{y^{(k)} + \varepsilon D_x\eta^{(k-1)} + O(\varepsilon^2)}{1 + \varepsilon D_x\xi + O(\varepsilon^2)} \tag{6}$$

$$= \left(y^{(k)} + \varepsilon D_x\eta^{(k-1)}\right)\left(1 - \varepsilon D_x\eta^{(k-1)}\right) + O(\varepsilon^2).$$

Thus,

$$\eta^{(k)}\left(x, y, y', \ldots, y^k\right) = \left(D_x\eta^{(k-1)} - y^{(k)}D_x\xi\right) \tag{7}$$

Mimicking the notation of chapter 1, we may write this in terms of the characteristic $Q = \eta - y'\xi$ by letting $\xi = -Q_y,\ \ \eta = Q - y'Q_{y'},\ \ \eta^{(k)} = D_x^k Q - y^{(k+1)}Q_{y'},\ \ k \geq 1.$

Recall that for first order ODEs the RHS of the linearized symmetry condition is $X\omega,\ X = \xi\partial_x + \eta\partial_y$, and the tangent vector to the orbit through (x, y) is $(\xi, \eta) = \left(\frac{d\hat{x}}{D\varepsilon}, \frac{d\hat{Y}}{D\varepsilon}\right)_{\varepsilon=0}$. For n^{th} order ODEs we have

$$X^{(n)} = \ \xi\partial_x + \eta\partial_y + \eta^{(1)}\partial_{y'} + \cdots + \eta^{(n)}\partial_{y^{(n)}} \tag{8}$$

defining the **prolonged infinitesimal generator**. Thus $X^{(n)}$ is associated with the tangent vector of the space variables $\left(x, y, y', \ldots, y^{(n)}\right)$. So far every symmetry we have met is a diffeomorphism of the form $(\hat{x}, \hat{y}) = \left(\hat{x}(x, y), \hat{y}(x, y)\right)$, which we call a ***point transformation***. Any point transformation that is a symmetry is ***point symmetry***. Let us stick with point symmetries until further notice. Then for point symmetries we have $\xi = \xi(x, y)$ and $\eta = \eta(x, y)$ only. These components of the tangent vector do not depend on $y', y'', \ldots$. Assuming that we have our symmetry $(\hat{x}, \hat{y}) = \left(\hat{x}(x, y), \hat{y}(x, y)\right)$ and the corresponding tangent vector $(\xi, \eta,)$ let's make the first few cases of equation (7) explicit, say for $k = 1{,}2{,}3$. For $k = 1$,

$$\eta^{(1)} = D_x\eta - y' D_x\xi \tag{9}$$

$$= \left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\eta$$

$$- y'\left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\xi$$

$$= \eta_x + y'\eta_y - y'\xi_x - \xi_y {y'}^2 = \eta_x + \left(\eta_y - \xi_x\right)y' - \xi_y {y'}^2$$

For $k = 2$,

$$\eta^{(2)} = D_x\eta^{(1)} - y'' D_x\xi \tag{10}$$

$$= \left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\left(\eta_x + \left(\eta_y - \xi_x\right)y' - \xi_y {y'}^2\right)$$

$$- y''\left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\xi$$

$$= \eta_{xx} + y'\eta_{xy} + {y'}^2\eta_{yy} + y'\eta_{xy} - y'\xi_{xx} - {y'}^2\xi_{xy} + \cdots$$

$$= \eta_{xx} + \left(2\eta_{xy} - \xi_{xx}\right)y' + \left(\eta_{yy} - 2\xi_{xy}\right){y'}^2 - \xi_{yy}{y'}^2$$

$$+ \left(\eta_y - 2\xi_x - 3\xi_y y'\right)y''.$$

For $k = 3$,

$$\eta^{(3)} = \eta_{xxx} + \left(3\eta_{xxy} - \xi_{xxx}\right)y' + 3\left(\eta_{xyy} - \xi_{xxy}\right){y'}^2 \tag{11}$$

$$+ \left(\eta_{yyy} - 3\xi_{xyy}\right){y'}^3 - \xi_{yyy}{y'}^4$$

$$+ 3\left(\eta_{xy} - \xi_{xx} + \left(\eta_{yy} - 3\xi_{xy}\right)y' - 2\xi_{yy}{y'}^2\right)y''$$

$$- 3\xi_y {y''}^2 + \left(\eta_y - 3\xi_x - 4\xi_y y'\right)y'''.$$

It gets increasingly monotonous. To keep things manageable and clear, let's consider second-order ODEs of the form $y'' = \omega(x, y, y')$. The linearized symmetry condition is obtained by substituting $\eta^{(1)}$ into $\eta^{(2)} \equiv \xi\omega_x + \eta\omega_y + \eta^{(1)}\omega_{y'}$, and then replacing $y''$by $\omega(x, y, y')$. We get

$$\eta^{(2)} = \eta_{xx} + \left(2\eta_{xy} - \xi_{xx}\right)y' + \left(\eta_{yy} - 2\xi_{xy}\right){y'}^2 - \xi_{yy}{y'}^3 + \left(\eta_y - 2\xi_x - 3\xi_y y'\right)\omega(x, y, y')$$

$$= \xi\omega_x + \eta\omega_y + \eta^{(1)}\omega_{y'} = \xi\omega_x + \eta\omega_y + \left(\eta_x + \left(\eta_y - \xi_x\right)y' - \xi_y {y'}^2\right)\omega_{y'}$$

Though the equation looks complicated, it is often easy to solve. As both ξ and η are independent of y', we may decompose the linearized symmetry condition into a system of partial differential equations (PDEs), which are the determining equations for the Lie point symmetries.

Example 2.2—Consider $y'' = 0$. Our goal is to find a Lie point symmetry. Yes we are going overboard with a very simple 2nd order ODE, but what we learn will **greatly simplify** not-so-simple 2nd order (and higher order) linear and nonlinear ODEs (**see next example**). Note $\omega = 0,\ \omega_x = 0,\ \omega_y = 0,\ \omega_{y'} = 0$. Thus $\eta^{(2)}$ simplifies to $\eta_{xx} + \left(2\eta_{xy} - \xi_{xx}\right)y' + \left(\eta_{yy} - 2\xi_{xy}\right){y'}^2 - \xi_{yy}{y'}^3 = 0$. As both ξ and η are independent of y', the linearized symmetry condition splits into the following system of determining equations:

$$\eta_{xx} = 0, \qquad 2\eta_{xy} - \xi_{xx} = 0, \qquad \eta_{yy} - 2\xi_{xy} = 0, \qquad \xi_{yy} = 0.$$

The general solution for $\xi_{yy} = 0$ is $\xi(x, y) = A(x)y + B(x)$. Take two partial derivatives wrt to y to verify that indeed $\xi_{yy} = 0$. Now consider the third equation $\eta_{yy} - 2\xi_{xy} = 0$. Then

$$\eta_{yy} = 2\xi_{xy} = 2\frac{\partial}{\partial x}\frac{\partial}{\partial y}[A(x)y + B(x)] = 2\frac{\partial}{\partial x}A(x) = 2A'(x).$$

Integrating twice we get $\eta = A'(x)y^2 + C(x)y + D(x)$. Then $\eta_x = A''(x)y^2 + C'(x)y + D'(x)$, and

$$\eta_{xx} = A'''(x)y^2 + C''(x)y + D''(x) = 0. \tag{12}$$

What is ξ_{xx}? First $\xi_x = A'(x)y + B'(x)$, then $\xi_{xx} = A''(x)y + B''(x)$. Then the second equation tells us $2\eta_{xy} = A''(x)y + B''(x)$. On the LHS, $2\eta_{xy} = 2 \cdot \left(2A''(x)y + C'(x)\right)$. So with LHS = RHS we get:

$$3A''(x)y + 2C'(x) - B''(x) = 0. \tag{13}$$

If $y \neq 0$ in (12), then $A'''(x) = 0$ and $C''(x) = 0$, and $D''(x) = 0$. Ditto if $y \neq 0$ in (13) then $A''(x) = 0$, and therefore $2C'(x) = B''(x)$. Integrating $A''(x) = 0$ gives us $A(x) = c_\alpha + c_\beta x$. Integrating $C''(x) = 0$ gives $C(x) = c_\delta + c_\gamma x$. Therefore, $B''(x) = 2C'(x) = 2c_\gamma$. Integrating $B''(x)$ once we get

$B'(x) = 2c_\gamma x + c_\theta$. One more integration gets us: $B(x) = c_\gamma x^2 + c_\Theta x + c_\rho$. Repeating for $D''(x) = 0$, we get $D(x) = c_\lambda x + c_\chi$. Thus our Lie point symmetry is

$$\xi(x,y) = A(x)y + B(x) = c_\alpha y + c_\beta xy + c_\gamma x^2 + c_\Theta x + c_\rho,$$

$$\eta(x,y) = A'(x)y^2 + C(x)y + D(x) = c_\beta y^2 + (c_\delta + c_\gamma x)y + c_\lambda x + c_\chi.$$

Let's re-label the Greek indices with Arabic numerals: Then $\xi(x,y) = c_1 + c_3 x + c_5 y + c_7 x^2 + c_8 xy$ and $\eta(x,y) = c_2 + c_4 y + c_6 x + +c_7 xy + c_8 y^2$. Now

$$X^{(n=0)} = \xi\partial_x + \eta\partial_y + \eta^{(1)}\partial_{y'} + \cdots + \eta^{(n)}\partial_{y^{(n)}} = \xi\partial_x + \eta\partial_y$$

$$= (c_1 + c_3 x + c_5 y + c_7 x^2 + c_8 xy)\partial_x + (c_2 + c_4 y + c_6 x + +c_7 xy + c_8 y^2)\partial_y.$$

The most general infinitesimal generator is $X = \sum_{i=1}^{8} c_i X_i$ where

$$X_1 = \partial_x, \quad X_2 = \partial_y, \quad X_3 = x\partial_x, \quad X_4 = y\partial_y, \quad X_5 = y\partial_x, \quad X_6 = x\partial_y,$$

$$X_7 = x^2\partial_x + xy\partial_y, \quad X_8 = xy\partial_x + y^2\partial_y$$

<u>Example 2.3</u>—Consider the nonlinear ODE $y'' = \frac{y'^2}{y} - y^2$. Recall the linearized symmetry condition

$$\eta^{(2)} = \eta_{xx} + (2\eta_{xy} - \xi_{xx})y' + (\eta_{yy} - 2\xi_{xy})y'^2 - \xi_{yy}y'^3 + (\eta_y - 2\xi_x - 3\xi_y y')\omega(x,y,y')$$

$$= \xi\omega_x + \eta\omega_y + \eta^{(1)}\omega_{y'} = \xi\omega_x + \eta\omega_y + (\eta_x + (\eta_y - \xi_x)y' - \xi_y y'^2)\omega_{y'}$$

The term $\omega(x,y,y')$ is the RHS of our nonlinear ODE. The term ω_x is zero. The term ω_y is $-\frac{y'^2}{y^2} - 2y$. The term $\omega_{y'}$ is $2y'/y$. Thus for our nonlinear ODE our linearized symmetry condition is

$$\eta_{xx} + (2\eta_{xy} - \xi_{xx})y' + (\eta_{yy} - 2\xi_{xy})y'^2 - \xi_{yy}y'^3 + \{\eta_y - 2\xi_x - 3\xi_y y'\}\left(\frac{y'^2}{y} - y^2\right)$$

$$= \eta\left(-\frac{y'^2}{y^2} - 2y\right) + (\eta_x + (\eta_y - \xi_x)y' - \xi_x y'^2)\left(\frac{2y'}{y}\right).$$

We proceed as before by matching powers. Matching powers of y' leads to the determining equations

$$\xi_{yy} + \frac{1}{y}\xi_y = 0, \tag{14}$$

$$\eta_{yy} - 2\xi_{xy} - \frac{1}{y}\eta_y + \frac{1}{y^2}\eta = 0, \tag{15}$$

$$2\eta_{xy} - \xi_{xx} + 3y^2\xi_y - \frac{2}{y}\eta_x = 0, \tag{16}$$

$$\eta_{xx} - y^2(\eta_y - 2\xi_x) + 2y\eta = 0. \tag{17}$$

Abusing notation, let's integrate the first of these equations

$$\int \frac{\partial \xi_y}{\xi_y} + \int \frac{\partial y}{y} = 0.$$

We get $\ln \xi_y + \ln y + \ln Z(x) = 0$, where $Z(x)$ is some arbitrary function of x. This can be rewritten as $\ln|\xi_y y Z(x)| = 0$. So, $\xi_y = 1/(Z(x)y)$. Integrating again gives us $\xi = (1/Z(x)) \cdot \ln|y| + B(x)$, where $B(x)$ is another arbitrary function of x. Let $1/Z(x) = A(x)$. Then

$$\xi = A(x)\ln|y| + B(x). \tag{18}$$

You can check this by taking ∂_{yy}. Integrating the second of the equations yields

$$\eta = A'(x)y(\ln|y|)^2 + C(x)y\ln|y| + D(x)y, \tag{19}$$

where, as usual, the functions of x are arbitrary. Substituting (18) and (19) into (16) results in

$$3A''(x)\ln|y| + 3A(x)y + 2C'(x) - B''(x) = 0. \tag{20}$$

If $y \neq 0$, then

$$A(x) = 0, \qquad B''(x) = 2C'(x). \tag{21}$$

Substituting (18) and (19) into (17) results in (remembering $A(x) = 0$.)

$$C(x)y^2\ln|y| + C''(x) + (2B'(x) - C(x) + D(x)y^2) + D''(x)y = 0,$$

which splits into the system

$$C(x) = 0, \qquad D(x) = -2B'(x), \qquad D''(x) = 0.$$

Taking into account equations (21), we see that

$$B(x) = c_1 + c_2 x, \qquad D(x) = -c_2,$$

where the constants c_1 and c_2 are arbitrary. Hence the general solution of the generalized symmetry condition is $\xi(\hat{x}, \hat{y}) = c_1 + c_2 x,\ \eta(\hat{x}, \hat{y}) = -2c_2 y$.

I'll bet that when I said symmetry methods for solving differential equations are based on making good guesses on symmetries, you thought that symmetry methods are just as full of crap as anything else. Maybe only a genius could cook up a symmetry for the Riccati equation. Not so. In this example we began with an ugly, nonlinear, 2nd order ODE and through the machinery of symmetry methods produced the tangent vector (ξ, η) to a symmetry of our ugly ODE. It may not be easy to work from the tangent vector back to the symmetry, but the tangent vector hasn't yet been fully exploited. Recall that our goal is to find solutions to our ODE.

Let's use our two infinitesimal generators to look directly for **possible** invariant solutions from the characteristic equation for each generator. Recall from chapter 1 that every curve C on the xy-plane that is invariant under the group generated by a particular X satisfies the characteristic equation: $Q(x, y, y') = \eta - y'\xi = 0$ on C. Let's try it for X_1. We know $X_1 = \partial_X$, and thus for $X_1,\ \xi = 1$ and $\eta = 0$. Thus for X_1 the characteristic equation is $0 - y' \cdot 1 = 0$. The solution to this characteristic equation is $y = c$, where c is an arbitrary constant. If you plug in this possible solution into the ugly ODE it works. So $y = c$ is a confirmed solution. Now for our second characteristic equation for X_2. Here $\xi = x$ and $\eta = -2y$, hence we determine when $-2y - y' \cdot x = 0$. The solution to this simple ODE is $y = y_0/x^2$ where y_0 is an arbitrary constant. Is this solution a solution to our nonlinear ODE? Plug in and check. First let's compute the derivatives: $y = \frac{y_0}{x^2}$, $y' = \frac{-2y_0}{x^3}$, $y'' = \frac{6y_0}{x^4}$. Plug these into

$y'' = \frac{y'^2}{y} - y^2$ to get

$$\frac{6y_0}{x^4} = \frac{4{y_0}^2}{x^6 y} - y^2 = \frac{4{y_0}^2}{x^6} \cdot \frac{x^2}{y_0} - \frac{{y_0}^2}{x^4} = \frac{4y_0}{x^4} - \frac{{y_0}^2}{x^4}.$$

This simplifies to $6 = 4 - y_0$, or $y_0 = -2$. Thus $y = -2/x^2$is a solution. Check it:

$$y'' = -\frac{12}{x^4} = \frac{y'^2}{y} - y^2 = \frac{16}{x^6} \cdot \left(-\frac{x^2}{2}\right) - \frac{4}{x^4} = -\frac{12}{x^4}.$$

Not bad. Be careful. For linear ODEs the sum of solutions is a solution. This is not generally so for nonlinear differential equations. We have found two distinct solutions: $y = c$ and $y = -2/x^2$.

Get ready for some excitement—the **useful** union of abstract algebra to differential equations. We will reduce our ODE down to an equivalent algebraic equation. Let's first check if the algebra of the infinitesimal generators of our ODE is solvable. That is, let's check to see whether $\{X_1, X_2\} \longmapsto \cdots \longmapsto \{0\}$ or not under commutation. Well,

$$[X_1, X_2] = \partial_x(x\partial_x - 2y\partial_y) - (x\partial_x - 2y\partial_y)\partial_x = \partial_x + x\partial_{xx} - 2y\partial_{xy} - x\partial_{xx} + 2y\partial_{xy} = \partial_x = X_1.$$

Since $[X_1, X_1] = 0$, $\{X_1, X_2\} \longmapsto \{X_1\} \longmapsto \{0\}$. To go from $\{X_1, X_2\}$ to $\{0\}$ requires two steps. Thus our 2nd order ODE can (at least in principle) be integrated twice via differential invariants/canonical coordinates. [**<u>I owe you a proof of this, and you will get it</u>**.] One technique for reducing the order of a differential equation is the reduction of order using canonical coordinates. We've used canonical coordinates already for solving an example in chapter 1. Another technique involves computing differential invariants (this approach will be presented after canonical coordinates). When the algebra is solvable, we get the best of these approaches.

Since we have used canonical coordinates before, let's use them to reduce the order of our 2nd order ODE down to first order. Actually, you'll see that we can reduce our particular 2nd order ODE to two distinct 1st order ODEs depending on our choice of canonical coordinates. Recall that we can use canonical coordinates when the ODE has Lie symmetries equivalent to a translation, *e.g.*, $(\hat{x}, \hat{y}) = (x, y + \varepsilon)$. For $X_1 = \partial_X$, we have $\xi = 1$ and $\eta = 0$, so (chapter 1) $s(r, x) = \left(\int \frac{dx}{\xi(x, y(r,x))}\right)_{r=r(x,y)} = \int \frac{dx}{1} = x$ and $\frac{dy}{dx} = \frac{\eta(x,y)}{\xi(x,y)} = \frac{0}{1}$. Thus $y = y_o = \text{constant} = r$. So our canonical coordinates are $(r, s) = (y, x)$. Then $\frac{ds}{dx} = 1$ and $\frac{dr}{dx} = y'$, and so $\frac{ds}{dr} = ds/dx \div dr/dx = \frac{1}{y'}$. Let $v = \frac{ds}{dr} = \frac{1}{y'}$. With $\frac{dv}{dx} = -\frac{y''}{y'^2}$ and $\frac{dr}{dx} = y'$ we get

$$\frac{dv}{dr} = \frac{\frac{dv}{dx}}{\frac{dr}{dx}} = \frac{d^2 s}{dr^2} = -\frac{y''}{y'^3}.$$

Dividing our ODE by y'^2 and rearranging terms we get to

$$-\frac{y''}{y'^3} = \frac{y^2}{y'^3} - \frac{1}{yy'}.$$

Combining the two result above gives us the following first order ODE:

$$\frac{dv}{dr} = \frac{y^2}{y'^3} - \frac{1}{yy'} = r^2 v^3 - \frac{v}{r}.$$

On the other hand, if we let $v = \frac{1}{\dot{s}} = y'$, then $\frac{dv}{dx} = y''$, and with $\frac{dr}{dx} = y'$ we get

$$\frac{dv}{dr} = \frac{y''}{y'} = \frac{y'}{y} - \frac{y^2}{y'} = \frac{v}{r} - \frac{r^2}{v}.$$

I suppose which choice of v is better is the one that produces the easier reduced ODE to work with. Regardless, we have reduced a 2nd order ODE into one or another first order ODE, but according to the fact that we have two Lie point symmetries, we can transform our 2nd order ODE down to an algebra equation. Although we will need to build more machinery to more systematically reduce ODEs to lower order you have enough background to peek ahead and see how to take our particular 2nd order ODE down to an algebra equation. You'll have to take one result on my word until we build up this math.

So far we have only used the first infinitesimal generator $X_1 = \partial_X$ of our 2nd order ODE. Let's proceed to use the second infinitesimal generator $X_2 = x\partial_X - 2y\partial_y$. To keep things clear we will use subscripts. Corresponding to generator X_1 we have canonical coordinates $(r_1, s_1) = (y, x)$. Using the subscript 2 for X_2, we have $\xi_2 = x$ and $\eta_2 = -2y$., and equation (8) reduces to

$$X^{(n)} = \xi_2\partial_x + \eta_2\partial_y + \eta_2^{(1)}\partial_{y'}. \tag{22}$$

The method of characteristics tells us

$$\frac{dx}{\xi_2} = \frac{dy}{\eta_2} = \frac{dy'}{\eta_2^{(1)}}. \tag{23}$$

What is $\eta_2^{(1)}$? Equation 9 tells us that

$$\eta_2^{(1)} = \eta_{2_x} + \left(\eta_{2_y} - \xi_{2_x}\right)y' - \xi_{2_y}y'^2.$$

This reduces to $\eta_2^{(1)} = (-2-1)y' = -3y'$. So equation (23) becomes

$$\frac{dx}{x} = -\frac{dy}{2y} = -\frac{dy'}{3y'}. \tag{24}$$

Let's work with the last two terms. Integrating the last two terms gives $y' = y^{\frac{3}{2}}$. Thus $y'y^{-\frac{3}{2}} = \text{constant} = r_2$. **Until I cover the material for differential invariants you will have to take my word that**: $v_2 = \frac{y''}{y^2}$ (see example 2.7). Our canonical coordinates are (r_2, v_2). Notice that if I rearrange the original ODE to match $\frac{y''}{y^2}$ we get $v_2 = \frac{y''}{y^2} = \frac{y'^2}{y^3} - 1 = \frac{y' \cdot y'}{y^{\frac{3}{2}} \cdot y^{\frac{3}{2}}} - 1 = r_2^2 - 1$, and there you have it. A second order ODE turned into an algebraic equation. **Since we're not interested in voodoo, here is what I owe you: the theory of differential invariants to get to $v_2 = \frac{y''}{y^2}$. I also owe you the connection between the solvability of the Lie algebra, $\{X_1, X_2\} \longmapsto \{X_1\} \longmapsto \{0\}$, and the ability to integrate our 2nd order ODE stepwise by two integrations. We will get here soon. We will do so for our particular example, taking it to completion, and we will do so in general for any ODE and its Lie algebra.**

It's remarkable how we've treated our 2nd order ODE by assuming that at least one Lie point symmetry exists and using the linearize symmetry conditions to get tangent vectors to the Lie symmetries, to get some invariant solutions, to get the infinitesimal generators X_1 and X_2, to get canonical coordinates, to reduce our ODE down to two different first order ODEs [no voodoo through here], and finally reduce the 2nd order ODE down to a polynomial. Let's do two more examples of reduction of order using canonical coordinates before formalizing the examples into our grammar afterwards.

<u>Example 2.4</u>—Consider the 2nd order ODE $y'' = \left(\frac{3}{x} - 2x\right) y' + 4y$.

Since I did every step in using the linearized symmetry condition, I will forego all of that here and state that one of the infinitesimal generators is $X = y\partial_y$. So for this generator $\xi = 0$ and $\eta = y$. Given this, we may derive the canonical coordinates $(r, s) = (x, \ln|y|)$. We have seen these before. Then we get

$$v = \frac{ds}{dr} = \frac{\frac{ds}{dx}}{\frac{dr}{dx}} = \frac{y'}{y}.$$

We are clearly aiming to reduce the order of our ODE from two to one. When we "prolong" v, we get

$$\frac{dv}{dr} = \frac{\frac{ds}{dx}}{\frac{dr}{dx}} = \frac{d^2 s}{dr^2} = \frac{y''}{y} - \frac{{y'}^2}{y^2}.$$

So

$$\frac{y''}{y} = \frac{dv}{dr} + \frac{{y'}^2}{y^2}.$$

If we divide our ODE by y, we get

$$\frac{y''}{y} = \left(\frac{3}{x} - 2x\right)\frac{y'}{y} + 4.$$

Combining the two above equations gets us

$$\frac{y''}{y} = \frac{dv}{dr} + \frac{{y'}^2}{y^2} = \frac{dv}{dr} + v^2 = \left(\frac{3}{r} - 2r\right)v + 4.$$

Thus, again, we arrive at a first order equation (in this case a Riccati equation):

$$\frac{dv}{dr} = \left(\frac{3}{r} - 2r\right)v + 4 - v^2.$$

Wikipedia writes a general Riccati equation as $y' = q_1(x)y^2 + q_1(x)y + q_0(x)$.

<u>Example 2.5</u>—Being a little adventuresome. Consider $y'' = \frac{{y'}^2}{y} + \left(y - \frac{1}{y}\right)y'$.

Again, exploiting the linearized symmetry condition, we can determine that one of the infinitesimal generators is $X = \partial_x$. (These translations are the only Lie point symmetries of our ODE.) Since $\xi = 1$ and η=0 we get canonical coordinates $(r, s) = (y, x)$. With $\frac{ds}{dx} = 1$ and $\frac{dr}{dx} = y'$, let $v = \frac{ds}{dr} = ds/dx \div dr/dx = \frac{1}{y'}$. With $\frac{dv}{dx} = -\frac{y''}{{y'}^2}$ and $\frac{dr}{dx} = y'$we get

$$\frac{dv}{dr} = \frac{\frac{dv}{dx}}{\frac{dr}{dx}} = \frac{d^2s}{dr^2} = -\frac{y''}{{y'}^3}.$$

Then

$$\frac{dv}{dr} = -\frac{y''}{y'^3} = \left(\frac{1}{y} - y\right)\frac{1}{y'^2} - \frac{1}{yy'} = \left(\frac{1}{r} - r\right)v^2 - \frac{v}{r}.$$

This is a Bernoulli equation whose solution is given in the typical botany-based class in ODEs. What if instead of choosing $v = \frac{ds}{dr} = \frac{1}{y'}$ we choose $v = y'$? Then $dv/dx = y''$ and $dr/dx = y'$. Then

$$\frac{dv}{dr} = \frac{\frac{dv}{dx}}{\frac{dr}{dx}} = \frac{y''}{y'} = \frac{y'}{y} + \left(y - \frac{1}{y}\right) = \frac{v}{r} + \left(r - \frac{1}{r}\right).$$

The result is a linear ODE, yet another of the ODEs studied in a regular, botany-based ODE class. ¿Botany? Ha! We don't need no stinckin' botany. Vamonos!

So far we have been able to reduce the order of our 2nd order ODEs by one because each of the these 2nd order ODEs have had at least one Lie point symmetry, and therefore one corresponding infinitesimal generator extracted from the linearized symmetry conditions. We don't need the actual Lie point symmetries because we extract what we need from the tangent vectors of these Lie point symmetries which we get from the linearized symmetry condition. Some of our ODEs have actually had two infinitesimal generators extracted from the linearized symmetry conditions, and in some cases these pairs of generators have formed a solvable Lie algebra. For these 2nd order ODEs with solvable algebras I have claimed that we can integrate them stepwise two times. We haven't bothered to do this for the prior two examples because the reduced 1st order ODEs we have arrived at are straightforward 1st order ODEs that we can readily solve. Let's formalize this machinery of reducing the order of an ODE by using canonical coordinates. Then we will go over an example reducing a 3rd order ODE down to a 1st order ODE. Ultimately we will use stronger machinery to more fully treat example 2.3. What follows is the formalization of the reduction of order technique we used in examples 2.4 and 2.5 using canonical coordinates.

Reduction of order by using canonical coordinates: From now on we shall denote differentiation by r by a dot $(\cdot)$. For example, $\dot{s}$ denotes ds/dr. Suppose that X is an infinitesimal generator of a one-parameter Lie group of symmetries of ODE

$$y^{(n)} = \omega\left(x, y, y', \dots, y^{(n-1)}\right), \qquad n \geq 2. \tag{22}$$

Let (r, s) be canonical coordinates for the group generated by X, so that $X = \partial_s$. If ODE (22) is written in terms of canonical coordinates, it is of the form

$$s^{(n)} = \Omega\left(r, s, \dot{s}, \dots, s^{(n-1)}\right), \qquad s^{(k)} = \frac{d^k s}{dr^k}. \tag{23}$$

for some Ω. However, ODE (23) is invariant under the Lie group of translations in s, so the symmetry condition gives $\Omega_s = 0$. Therefore

$$s^{(n)} = \Omega\left(r, \dot{s}, \dots, s^{(n-1)}\right), \qquad s^{(k)} = \frac{d^k s}{dr^k}. \tag{24}$$

By writing ODE (22) in terms of canonical coordinates, we have reduced it to an ODE of order $n-1$ for $v = \dot{s}$. This is just what I did in example 2.3 just south of equation (24). So with $v = \dot{s}$,

$$v^{(n-1)} = \Omega\left(r, v, \dots, v^{(n-2)}\right), \qquad v^{(k)} = \frac{d^{k+1} s}{dr^{k+1}}. \tag{25}$$

Suppose (assume) that the reduced ODE has general solution

$$v = f(r; c_1, \dots, c_{n-1}).$$

Then the general solution of ODE (22) is arrived at by integration,

$$s(x, y) = \int^{r(x,y)} f(r; c_1, \dots, c_{n-1}) dr + c_n.$$

More generally, if v is any function of $\{\dot{s}\,, r\}$ so that $v_x(r, \dot{s}\,) \neq 0$, then(24) reduces to form

$$v^{(n-1)} = \widetilde{\Omega}\big(r, v, \dots, v^{(n-2)}\big), \qquad v^{(k)} = \frac{d^k v}{dr^k}. \tag{26}$$

Once the solution of (26) is known, the relationship $s = \dot{s}(\dot{r}, v)$ gives the general solution of (22):

$$s(x, y) = \int^{r(x,y)} \dot{s}(r; v(r; c_1, \dots, c_{n-1}) dr + c_n. \tag{27}$$

In summary, once we find a one-parameter Lie group of symmetries using whatever means (say a good guess) or the tangent vectors to any such one-parameter Lie group via the linearized symmetry conditions), we can solve ODE (22) by solving a lower-order ODE, then integrating. This is what we have done in examples 2.3 through 2.5.

On Differential Invariants and Reduction of Order: In just about as much theory of the previous section we will have the differential invariants method to reduce the order of a differential equation. We have seen that a single Lie point symmetry allows us to reduce the order of an ODE by one. Sometimes a double reduction of order can happen if there are two Lie point symmetries. In fact one can reduce an n^{th} order ODE with $R \leq n$ Lie point symmetries to an ODE of order $n - R$, or to an algebraic equation if $R = n$.

If X generates Lie point symmetries of the ODE

$$y^{(n)} = \omega\big(x, y, y', \dots, y^{(n-1)}\big), \qquad n \geq 2 \tag{28}$$

then, in terms of canonical coordinates $(r, \dot{s})$, the ODE reduces to

$$v^{(n-1)} = \Omega\big(r, v, \dots, v^{(n-2)}\big), \tag{29}$$

Where $\nu = \nu(r, \dot{s})$ is any function such that $\nu_{\dot{s}} \neq 0$. The reduced ODE consists entirely of functions that are invariant under the (prolonged) action of the group generated by $X = \partial_s$. These functions are called ***differential invariants***. A nonconstant function $I\big(x, y, y', \dots, y^{(k)}\big)$ is a k^{th} order differential invariant of the group generated by X if

$$X^{(k)}I = 0. \tag{29}$$

In canonical coordinates, $X^{(k)} = \partial_s$, so that every k^{th} order differential invariant is of the form $I = F(r, \dot{s}, \dots, s^{(k)})$, or equivalently,

$$I = F(r, v, \dots, v^{(k-1)}) \tag{30}$$

for some function F. The invariant canonical coordinate $r(x, y)$is the only differential invariant of order zero (up to function dependence). All first-order differential invariants are functions of $r(x, y, y')$ and $v(x, y, y')$. Furthermore, all differential invariants of order two or greater are functions of r, v and derivatives of v wrt r. Therefore, r and v are called *fundamental differential* invariants. We can usually find a pair of fundamental differential invariants without first having to determine s. From equation (29), every k^{th} order differential invariant satisfies $\xi I_x + \eta I_y + \cdots + \eta^{(k)} I_{y^{(k)}} = 0$,

so (by the method of characteristics), I is a first integral of

$$\frac{dx}{\xi} = \frac{dy}{\eta} = \cdots = \frac{dy^{(k)}}{\eta^{(k)}}. \tag{31}$$

In particular, r is a first integral of $\frac{dx}{\xi} = \frac{dy}{\eta}$, and $v$is a first integral of $\frac{dx}{\xi} = \frac{dy}{\eta} = \frac{dy'}{\eta^{(1)}}$. Sometimes it is necessary to use r to obtain v.

<u>Example 2.6</u>—Find fundamental differential invariants of the group of rotations generated by $X = -y\partial_x + x\partial_y$. Thus $\xi = -y$ and $\eta = x$. Then $\frac{dy}{dx} = \frac{\eta}{\xi} = -\frac{x}{y}$. Solving this simple ODE leads to $x^2 + y^2 = \text{constant} = r^2$. I let the constant be r^2 because the left hand side is the equation of a circle. Then $r = \sqrt{x^2 + y^2}$. Note that $Xr = (-y\partial_x + x\partial_y)\sqrt{x^2 + y^2} = 0$. That is $Xr = 0$. We already know that r is a first integral of $\frac{dy}{dx} = -\frac{x}{y}$. Now on to v. It is the first integral of $-\frac{dx}{y} = \frac{dy}{x} = \frac{dy'}{\eta^{(1)}}$. Well $\eta^{(1)} = \eta_x + (\eta_y - \xi_x)y' - \xi_y y'^2 = 1 + y'^2$. So v is a first integral of

$$-\frac{dx}{y} = \frac{dy}{x} = \frac{dy'}{1+y'^2}.$$

To keep things simple we restrict attention to the region $x > 0$, where $\frac{dy}{x} = \frac{dy}{\sqrt{r^2-y^2}}$. Note that

$\int \frac{dy'}{1+y'^2} = \tan^{-1} y'$ and $\int \frac{dy'}{1+y'^2} = \int \frac{dy}{\sqrt{r^2-y^2}} = \sin^{-1}\frac{y}{r} = \tan^{-1}\frac{y}{x}$. So $\tan^{-1} y' - \tan^{-1}\frac{y}{x} = \text{constant}$.

Then the first integrals are of the form $I = F\left(r, \tan^{-1} y' - \tan^{-1}\frac{y}{x}\right)$. Since $\tan^{-1} y' - \tan^{-1}\frac{y}{x}$ is constant so is $\tan\left(\tan^{-1} y' - \tan^{-1}\frac{y}{x}\right) = \frac{xy'-y}{x+yy'}$ by identity. Then it's convenient to let $v = \frac{xy'-y}{x+yy'}$.

What is s? By definition,

$$s(r,x) = \left(\int \frac{dx}{\varepsilon\big(x, y(r,x)\big)}\right)_{r=r(x,y)} = -\int \frac{dx}{y} = -\int \frac{dx}{\sqrt{r^{2-x^2}}} = \sin^{-1}\frac{x}{r} = \tan^{-1}\frac{y}{x}.$$

Let's compute $\dot{s}$, or $\frac{ds}{dr}$.

$$\dot{s} = \frac{ds}{dr} = \frac{\frac{ds}{dx}}{\frac{dr}{dx}} = \frac{\frac{1}{1+\frac{y^2}{x^2}} \cdot \left(\frac{y'}{x} - \frac{y}{x^2}\right)}{\frac{-(x+yy')}{(x^2+y^2)^{1/2}}} = -\frac{1}{\sqrt{x^2+y^2}} \cdot \frac{(xy'-y)}{(x+yy')}.$$

Notice that $v = r\dot{s}$. We can think of $\dot{s}$ as angular velocity and $v$as the tangential velocity.

An ODE that has more than one Lie point symmetry can be written in terms of the differential invariants of each generator. Thus the generators themselves can be written in terms of functions that are invariant under all of its symmetries. Let $\mathcal{L}$ denote the set of all infinitesimal generators of one-parameter Lie groups (the binary operation being commutation) of point symmetries of an ODE of order $n \geq 2$. The linearized symmetry condition is linear in ξ and η by construction, and so

$$X_{1,} X_2 \in \mathcal{L} \Longrightarrow c_1 X_1 + c_2 X_{2,} \forall\, c_1, c_2 \in \mathbb{R},$$

Where $\{X_{1,} \dots, X_R\}$ is a **basis** for $\mathcal{L}$. The set of point symmetries generated by all $X \epsilon \mathcal{L}$ forms an R-parameter (local) Lie group called the group generated by $\mathcal{L}$. (The order of an ODE places restrictions on R. For 2nd order ODEs, $R = 0,1,2,3, \text{or } 8$. $R = 8$ iff the ODE is linear, or is linearizable by a point transformation. Every ODE of order $n \geq 3$ has $R \leq n + 4$. If this ODE is linear or linearizable, then $R = \{n + 1, n + 2, n + 4\}$. These results have been offered without proof.)

Because $\{X_{1,} \dots, X_R\}$ is a basis for $\mathcal{L}$, the fundamental differential invariants of the group generated by $\mathcal{L}$ are solutions of the system

$$\begin{bmatrix} \xi_1 & \eta_1 & \eta_1^{(1)} & \cdots & \eta_1^{(R)} \\ \xi_2 & \eta_2 & \eta_2^{(1)} & & \eta_2^{(R)} \\ & \vdots & & \ddots & \vdots \\ \xi_R & \eta_R & \eta_R^{(1)} & \cdots & \eta_R^{(R)} \end{bmatrix} \begin{bmatrix} I_x \\ I_y \\ I_{y'} \\ \vdots \\ I_{y^{(R)}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{32}$$

(A good example is coming up.) The system has two functionally independent solutions if the matrix on the LHS has rank R. These solutions can be determined via Gaussian elimination and the methods of characteristics. One solution is independent of $y^{(R)}$ and is denoted by r_R. We denote the other solution v_R, which depends nontrivially on $y^{(R)}$. As ${dv_r}/{dr_R}$ depends on $y^{(R+1)}$, and so on, ODE (22) reduces to

$$v_R^{(n-R)} = \Omega\left(r_R, v_R, \dots, v_R^{(n-R-1)}\right), \qquad v_R^{(k)} = \frac{d^k v_R}{dr_R^k}$$

for some function Ω. Thus an R- parameter symmetry group enables us to reduce the order of the ODE by R.

<u>**Example 2.7**</u>—Consider $y^{(iv)} = \frac{2}{y}(1 - y')y'''$. Is this ugly enough for you? As in example 2.3, you would apply the linearized symmetry condition to find $\xi, \eta, \eta^{(1)}, \dots, \eta^{(iv)}$. From these we would find the infinitesimal generators. Recall example 2.3. There we got $\xi(\hat{x}, \hat{y}) = c_1 + c_2 x, \text{ and } \eta(\hat{x}, \hat{y}) = -2c_2 y$.

Thus there were two constants in that example, and (if you go back) we had two infinitesimal generators, namely, $X_1 = \partial_X$ and $X_2 = x\partial_X - 2y\partial_y$. Doing this math for our 4th order ODE you get:

$$X_1 = \partial_x, \qquad X_2 = x\partial_x + y\partial_y, \qquad X_3 = x^2\partial_x + 2xy\partial_y. \tag{33}$$

So $R = 3$, and for X_1: $\xi_1 = 1,\ \eta_1 = 0,\ \eta_1^{(1)} = 0,\ \eta_1^{(2)} = 0,\ \eta_1^{(3)} = 0$. For X_2: $\xi_2 = x,\ \eta_2 = y,\ \eta_2^{(1)} = 0,\ \eta_2^{(2)} = -y'',\ \eta_2^{(3)} = -2y'''$. The result for $\eta_2^{(2)}$ follows from substituting ξ_2 and η_2 into (9):

$$\eta_2^{(1)} = \eta_{2_x} + \left(\eta_{2_y} - \xi_{2_x}\right)y' - \xi_{2_y}{y'}^2 = 0 + (1-1)\cdot y' - 0 = 0.$$

The result for $\eta_2^{(2)}$ results from substituting $\xi_2,\ \eta_2,$ and $\eta_2^{(1)}$ into equation (10)

$$\eta_2^{(2)} = \eta_{2_{xx}} + \left(2\eta_{2_{xy}} - \xi_{2_{xx}}\right)y' + \left(\eta_{2_{yy}} - 2\xi_{2_{xy}}\right){y'}^2 - \xi_{2_{yy}}{y'}^2 + \left(\eta_{2_y} - 2\xi_{2_x} - 3\xi_{2_y}y'\right)y''$$
$$= 0 + (0-0)\cdot y' + (0-0)\cdot {y'}^2 + (1-2-0)\cdot y'' = -y''.$$

The result for $\eta_2^{(3)}$ results from substituting $\xi_2,\ \eta_2,\ \eta_2^{(1)}$ into equation (11) to get (you do the math) $\eta_2^{(3)} = -2y'''$. For the third index $\xi_3 = x^2$ and $\eta_3 = 2xy$. Repeating the use of equations (9), (10) and (11) gives us $\eta_3^{(1)} = 2y,\ \ \eta_3^{(2)} = 2(y' - xy'')$ and $\eta_3^{(3)} = -4xy'''$. Thus we get

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ x & y & 0 & -y'' & -2y''' \\ x^2 & 2xy & 2y & 2(y'-xy'') & -4xy''' \end{bmatrix} \begin{bmatrix} I_x \\ I_y \\ I_{y'} \\ I_{y''} \\ I_{y'''} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

After Gaussian elimination we get

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & y & 0 & -y'' & -2y''' \\ 0 & 0 & y & y' & 0 \end{bmatrix} \begin{bmatrix} I_x \\ I_y \\ I_{y'} \\ I_{y''} \\ I_{y'''} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{34}$$

So $I = I(x, y, y', y''y''')$. Now use each equation of (34) to determine the differential invariants. The third equation is

$$yI_{y'} + y'I_{y''} = 0.$$

Using the method of characteristics we get $\frac{dy'}{y} = \frac{dy''}{y'}$. Solving this leads to $\int y'\,dy' = y\int dy''$, or, $\frac{1}{2}y'^2 = yy'' + c$. Since ${y'}^2 - 2yy''$ is constant, it is one of our differential invariants. So

$$I = I\left(x, y, 2yy'' - {y'}^2, y'''\right)$$

The second equation from (34) is

$$yI_y - y''I_{y''} - 2y''I_{y'''} = 0.$$

The method of characteristics gives us $\frac{dy}{y} = -\frac{dy''}{y''} = -\frac{dy'''}{2y'''}$. Integrating the first and third terms $\ln y = -\frac{1}{2}\ln y'''$, or $y''' = c/y^2$. So y^2y''' is constant.

$$I = I(x, y, r_3, v_3) = I\left(x, 2yy'' - {y'}^2, y^2y'''\right).$$

We've eliminated y. The third equation is $I_x = 0$, or via the method of characteristics $dx/x = 0$. Then $x = c$. Therefore $I = I\left(2yy'' - {y'}^2, y^2y'''\right)$. So the fundamental differential invariants of the group generated by our generators (33) are $r_3 = 2yy'' - {y'}^2$ and $v_3 = y^2y'''$. Higher order differential invariants can now be computed, say dv_3/dr_3.

$$\frac{dv_3}{dr_3} = \frac{D_xv_3}{D_xr_3} = \frac{\left(\partial_x + y'\partial_y + y''\partial_{y'+} + y'''\partial_{y''+} + y^{(iv)}\partial_{y'''} + \cdots\right)v_3}{\left(\partial_x + y'\partial_y + y''\partial_{y'+} + y'''\partial_{y''+} + y^{(iv)}\partial_{y'''} + \cdots\right)r_3} = \frac{2yy'y''' + y^2y^{(iv)}}{2y'y'' - 2y'y'' + 2yy'''}$$

$$= \frac{2y'y''' + yy^{(iv)}}{2y'''} = \frac{yy^{(iv)}}{2y'''} + y' = 1.$$

Note that our ODE $y^{(iv)} = \frac{2}{y}(1-y')y'''$, when rearranged is $\frac{yy^{(iv)}}{2y'''} + y' = 1 = \text{constant}$(a differential invariant). Solving $\frac{dv_3}{dr_3} = 1$ gives us

$$v_3 = r_3 + c_1.$$

The ODE has been turned in the algebra equation $v_3 = r_3 + c_1$ equivalent to a 3rd order ODE, namely,

$$y''' = \frac{2yy'' - y'^2 + c_1}{y^2}.$$

This 3rd order ODE is invariant under the three parameter Lie group generated by $\mathcal{L}$. By the way, I have now showed you the method by which I got $v_2 = \frac{y''}{y^2}$ back in example 2.3.

Of course we can proceed backwards. A set of fundamental differential invariants can be used to construct ODEs that have given Lie point symmetries. If (r_R, v_R) are fundamental differential invariants of an R- dimensional Lie group G, then every ODE (28) of order $n \geq R$ that has G as its symmetry group can be written in the form

$$v_R^{(n-R)} = \Omega\left(r_R, v_R, \dots, v_R^{(n-R-1)}\right), \tag{35}$$

for some function F. By expressing (35) in terms of $\left(x, y, \dots y^{(n)}\right)$, one obtains a family of ODEs that have the desired symmetries. Some of these ODEs may have extra symmetries. Let's quickly see how this "going backwards" idea works. After this we finally get to the solvability theory.

Example 2.8—Suppose that the fundamental differential invariants of the three-parameter group generated by $X_1 = \partial_x,\ X_2 = \partial_y,\ X_3 = x\partial_x + y\partial_y$ are $r_3 = y',\ v_3 = y'''/y''^2$. The most general third-order ODE with these symmetries is $v_3 = F(r_3)$, which is equivalent to $y''' = y''^2 F(y')$. The most general fourth-order ODE with these symmetries is equivalent to $dv_3/dr_3 = F(r_3, v_3)$. First let's compute dr_3/dx. We get $dr_3/dx = y''$. Now let's compute dv_3/dx. We get

$$\frac{dv_3}{dx} = \frac{y^{(iv)}y''^2 - 2y'''^2 y''}{y''^4}.$$

So then

$$\frac{\frac{dv_3}{dx}}{\frac{dr_3}{dx}} = \frac{dv_3}{dr_3} = \frac{y^{(iv)}y''^2 - 2y'''^2 y''}{y''^5} = \frac{y^{(iv)}}{y''^3} - \frac{2y'''^2}{y''^4} = F\left(y', \frac{y'''}{y''^2}\right).$$

Thus

$$y^{(iv)} = \frac{2y'''^2}{y''} + y''^3 F(r_3, v_3).$$

As they say in today's Air Force, the bottom line up front (BLUF) is: If an ODE has $R \leq n$ Lie point symmetries, we can reduce its order from n to $n - R$. **If $\boldsymbol{R = n}$, *<u>and</u>* if the algebra $\boldsymbol{\mathcal{L}}$ is solvable then (at least in principle if the manipulations are doable) we may, <u>additionally</u>, integrate the ODE $\boldsymbol{R}$ times to solve it.** This is a theorem. It requires proof, which requires some more foundations. **<u>Please skim</u>** the rest of this chapter at a minimum for now if you've just finished a sophomore course in ODEs and you wish to keep going with applications to ODEs and PDEs in chapter 3. The rest of this material is **<u>IMPORTANT</u>** if you are beginning graduate studies in physics so that you better understand the mathematical engine underlying quantum physics, or if you're either planning or are already doing work in particles and fields, even as a post-doctoral research fellow. The material here on the classification of symmetries is well developed via realizations versus the pedagogical half picture via matrix representations. Of course I will include examples, and I will try to highlight the important stuff in **<u>bold</u>**. **<u>DON'T</u> let the upcoming theory derail you from the momentum you have in applying symmetry methods for ODEs**. **<u>DON'T</u> worry about totally mastering the following theoretical material until you have the practical methods down pat**. **Lie point symmetries <u>ARE NOT</u> the only symmetries we can exploit regarding differential equations**. We'll meet some of these other symmetries in this chapter.

<u>Proof Section to Theorem (Some grammar/theory first)</u>—If $R = n$, and if the algebra $\mathcal{L}$ is solvable then (at least in principle) we may integrate the ODE R times to solve our ODE. Let's suppose that the Lie point symmetries of ODE

$$y^{(n)} = \omega\left(x, y, y', \dots, y^{(n-1)}\right), \qquad n \geq 2 \tag{1}$$

are generated by $\mathcal{L}$, which is R-dimensional. We know (and we have seen from several examples) that we may rewrite ODE (1) in terms of differential invariants (r_R, v_R) to obtain an ODE reduced down to order $n - R$, leaving us with an algebraic equation,

$$v_R = F(r_R; c_1, \dots, c_{n-R}),$$

which is equivalent to an ODE of order R with the R-parameter group of symmetries generated by $\mathcal{L}$. To answer our theorem we need to learn more about the structure of $\mathcal{L}$ (**stuff that is very important to graduate students of <u>physics</u>**).

Suppose that $X_1, X_2 \in \mathcal{L}$, where

$$X_i = \xi_i(x, y)\partial_x + \eta_i(x, y)\partial_y, \qquad i = 1{,}2. \tag{36}$$

The product $X_1 X_2$ is a second-order partial differential operator:

$$X_1 X_2 = \xi_1 \xi_2 {\partial_x}^2 + (\xi_1 \eta_2 + \eta_1 \xi_2)\partial_x \partial_y + \eta_1 \eta_2 {\partial_y}^2 + (X_1 \xi_2)\partial_x + (X_1 \eta_2)\partial_y,$$

where the last two terms have come from the chain rule. The product $X_2 X_1$ has the same 2nd order terms as $X_1 X_2$, but first order terms (from the chain rule) of $(X_2 \xi_1)\partial_x + (X_2 \eta_1)\partial_y$. So the commutator

$$\begin{aligned} [X_1, X_2] &= X_1 X_2 - X_2 X_1 \\ &= (X_1 \xi_2 - X_2 \xi_1)\partial_x + (X_1 \eta_2 - X_2 \eta_1)\partial_y. \end{aligned} \tag{37}$$

The commutator has many useful properties. Here are a few. It is ***antisymmetric***: $[X_1, X_2] = -[X_2, X_1]$. The commutator satisfies the ***Jacobi identity***

$$\left[X_1,[X_2,X_3]\right]+\left[X_2,[X_3,X_1]\right]+\left[X_3,[X_1,X_2]\right]=0.$$

The commutator is ***bilinear***, *i.e.*, linear in both arguments:

$$[c_1X_1+c_2X_2,X_3]=c_1[X_1,X_3]+c_2[X_2,X_3],$$

$$[X_1,c_2X_2+c_3X_3]=c_2[X_1,X_2]+c_3[X_1,X_3],$$

where the c_i are arbitrary constants.

The Lie algebra is invariant under a change of variables. Let's check this. Under a change of coordinates from (x,y) to (u,v), each generator X_i transforms according to the chain rule. Let's find out how this affects the commutator. Let $\check{X}_3=\left[\check{X}_1,\check{X}_2\right]$; the new accent mark is being used for writing the various X_i in terms of the new coordinates (u,v). That is in new coordinates (u,v) we have

$$\check{X}_i=(X_iu)\partial_u+(X_iv)\partial_v.$$

Let $F(u,v)$ be an arbitrary function. Then it must be that

$$\begin{aligned}\left[\check{X}_1,\check{X}_2\right]F &= \check{X}_1\{(X_2u)F_u+(X_2v)F_v\}-\check{X}_2\{(X_1u)F_u+(X_1v)F_v\}\\ &=(X_1X_2u)F_u+(X_1X_2v)F_v-(X_2X_1u)F_u-(X_2X_1v)F_v.=([X_1,X_2]u)F_u+([X_1,X_2]v)F_v\\ &=(X_3u)F_u+(X_3v)F_v.\end{aligned}$$

However, since F is an arbitrary function (meaning the results don't depend on F) we conclude that

$$\left[\check{X}_1,\check{X}_2\right]=(X_3u)\partial_u+(X_3v)\partial_v=\check{X}_3.$$

The commutator is therefore independent of the coordinate system in which it is being computed. We don't have to distinguish between $\check{X}_i$ and X_i.

All of the commuter properties that we have discussed so far have been in terms of commutators acting on the plane. The commutators of the prolonged generators

$$X_i^{(k)} = \xi\partial_x + \eta\partial_y + \eta^{(1)}\partial_{y'} + \cdots + \eta^{(k)}\partial_{y^{(k)}}$$

Is defined similarly: $\left[X_1^{(k)}, X_2^{(k)}\right] = X_1^{(k)}X_2^{(k)} - X_2^{(k)}X_1^{(k)}$.

Without loss of generality suppose, for example, that $X_1 = \partial_y,\ X_2 = \xi(x,y)\partial_x + \eta(x,y)\partial_y$. (Recall that our choice of coordinates does not affect the commutator.) Then

$$[X_1, X_2] = \partial_y\big(\xi(x,y)\partial_x + \eta(x,y)\partial_y\big) - \big(\xi(x,y)\partial_x + \eta(x,y)\partial_y\big)\partial_y = \xi_y\partial_x + \eta_y\partial_y = X_3.$$

From the beginning of the chapter, $\eta^{(1)} = (D_x\eta - y'D_x\xi)$, the prolongation formula for $X_3^{(1)}$is

$$X_3^{(1)} = \xi_y\partial_x + \eta_y\partial_y + \eta^{(1)}\partial_{y'}.$$

Plugging in to $\left[X_1^{(3)}, X_2^{(3)}\right]$ we get

$$\left[X_1^{(3)}, X_2^{(3)}\right] = \left[\partial_y, \xi_y\partial_x + \eta_y\partial_y + \eta^{(1)}\partial_{y'}\right] = \partial_y\big(\xi\partial_x + \eta\partial_y + \eta^{(1)}\partial_{y'}\big) - \big(\xi\partial_x + \eta\partial_y + \eta^{(1)}\partial_{y'}\big)\partial_y$$

$$= \xi_y\partial_x + \xi\partial_{xy} + \eta_y\partial_y + \eta\partial_{yy} + \eta_y^{(1)}\partial_{y'} - \xi\partial_{xy} - \eta\partial_{yy} = \xi_y\partial_x + \eta_y\partial_y + \eta_y^{(1)}\partial_{y'}$$

$$= X_3^{(1)}.$$

(In a different approach we could have used the fact that ∂_y and D_x commute ($D_x\partial_y = \big(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\big)\partial_y = \partial_y\big(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\big) = \partial_y D_x$.) Thus $[X_1, X_2] = X_3$ and $\left[X_1^{(3)}, X_2^{(3)}\right] = X_3^{(1)}$. The result continues for all $k \geq 1$. For $k > 1$ note that $D_x\eta_y^{(k-1)} - y^{(k)}D_x = \partial_y\big(D_x\eta^{(k-1)} - y^{(k)}D_x\xi\big) = \eta_y^{(k)}$, these relationships having been developed in earlier parts of this chapter. Thus $\left[X_1^{(k)}, X_2^{(k)}\right] = X_3^{(k)}\forall\ k$ in any coordinate system.

Now we show that if $X_i, X_j \epsilon\ \mathcal{L}$ then $\left[X_i, X_j\right] \in\ \mathcal{L}$, otherwise known as ***closure***. With $X_i^{(n)} = \xi_i\partial_x + \eta_i\partial_y + {\eta_i}^{(1)}\partial_{y'} + \cdots + {\eta_i}^{(n)}\partial_{y^{(n)}}$, we have that $X_i^{(n)}y^{(n)} = {\eta_i}^{(n)}$, all other terms but the last term being zero. Then $X_i^{(n)}\left(y^{(n)}-\omega\right)={\eta_i}^{(n)}-X_i^{(n)}\omega=0$ when $y^{(n)}-\omega=0$. For $n\geq 2$, the prolongation formula implies that ${\eta_i}^{(n)}$is linear in the highest derivative, $y^{(n)}$, whereas ω and therefore $X_i^{(n)}\omega$ are independent of the $y^{(n)}$. The linearized symmetry condition thus implies that $X_i^{(n)}\left(y^{(n)}-\omega\right)=\lambda_i\left(y^{(n)}-\omega\right)$ where $\lambda_i\left(x,y,y',\dots,y^{(n-1)}\right)=\frac{\partial\eta_i^{(n)}}{\partial y^{(n)}}$. Let $X=[X_1,X_2]$. Now consider $X^{(n)}\left(y^{(n)}-\omega\right)=\left[X_1^{(n)},X_2^{(n)}\right]\left(y^{(n)}-\omega\right)=X_1^{(n)}X_2^{(n)}\left(y^{(n)}-\omega\right)-X_2^{(n)}X_1^{(n)}\left(y^{(n)}-\omega\right)=X_1^{(n)}\left(X_2^{(n)}\left(y^{(n)}-\omega\right)\right)-X_2^{(n)}\left(X_1^{(n)}\left(y^{(n)}-\omega\right)\right)=X_1^{(n)}\lambda_2\left(y^{(n)}-\omega\right)-X_2^{(n)}\lambda_1\left(y^{(n)}-\omega\right)=\left(X_1^{(n)}\lambda_2-\lambda_1X_2^{(n)}\right)\left(y^{(n)}-\omega\right)$. Hence $X^{(n)}\left(y^{(n)}-\omega\right)=0$ when $y^{(n)}-\omega=0$, and therefore X generates Lie point symmetries. The same holds for the case for $n=1$, but it's complicated by the fact that $\eta_i^{(1)}$ is quadratic in y'. Recall: $\eta^{(1)}=\eta_x+\left(\eta_y-\xi_x\right)y'-\xi_y y'^2$. So, if $X_i,X_j\epsilon\,\mathcal{L}$ then $\left[X_i,X_j\right]\in\ \mathcal{L}$ (closure).

Important (physics) definitions. For $n\geq 2$, the set $\mathcal{L}$ is a finite-dimensional vector space. Once we pick a basis $\{X_1,\dots,X_R\}$ for $\mathcal{L}$, every generator of Lie point symmetries can be written as a linear combination of the generators in the chosen basis: $\left[X_i,X_j\right]=c_{ij}^kX_k$, where the c_{ij}^k are called the **structure constants**. If $\left[X_i,X_j\right]=0$, the generators $X_i$and $X_i$are said to commute. Clearly, every generator commutes with itself. Particle physicists deal with structure constants all the time, but they deal primarily with matrix representations of a Lie algebra, not with realizations of the infinitesimal generators as we're doing. Now let's get this out of the way. A vector space is defined to have certain properties which you can readily look up. If a vector space also has a product (in this case our commutator "product" $\left[X_i,X_j\right]=c_{ij}^kX_k$, then this vector space is a **Lie algebra**. A Lie algebra is a vector space that is closed under commutation where commutation is bilinear, antisymmetric, and satisfies the Jacobi identity. Antisymmetric means $\left[X_i,X_j\right]=-\left[X_j,X_i\right]$ which implies $c_{ij}^k=-c_{ji}^k$. Evidently $c_{ii}^k=0$ so we only compute the generators with $i<j$. The Jacobi identity $\left[X_1,[X_2,X_3]\right]+\left[X_2,[X_3,X_1]\right]+\left[X_3,[X_1,X_2]\right]=0\ \forall\ i,j,k$. It is true iff $c_{ij}^qc_{kq}^l+c_{jk}^qc_{iq}^l+c_{kl}^qc_{jq}^l=0\ \ \forall\ i,j,k.$

Example 2.9—The Lie algebra of Lie point symmetries of $y''' = y^{-3}$ is two-dimensional, being spanned by the basis $X_1 = \partial_x$, $X_2 = x\partial_x + \frac{3}{4}y\partial_y$. Then $[X_1, X_2] = \partial_x\left(x\partial_x + \frac{3}{4}y\partial_y\right) - \left(x\partial_x + \frac{3}{4}y\partial_y\right)\partial_x = \partial_x + x\partial_{xx} + \frac{3}{4}y\partial_{xy} - x\partial_{xx} - \frac{3}{4}y\partial_{xy} = \partial_x = X_1$. Now $[X_1, X_1]$=$[X_2, X_2] = 0$, and $[X_2, X_1] = -X_1$. Thus the only nonzero structure constants for the basis are $c_{12}^1 = 1$ and $c_{21}^1 = -1$.

Example 2.10—In the last example of chapter 1 (Example 1.15) I introduced you to the three-dimensional rotation matrices defining the group $SO(3)$, the rotation matrices providing the matrix representation of the $SO(3)$. The three matrices form the basis of $SO(3)$. The realization of $SO(3)$ came from the basis of quantum mechanical angular momentum operators derived from converting their classical mechanics versions to their quantum mechanical versions. **The basis for a representation or realization of $\boldsymbol{SO(3)}$ are not unique**. Here is another representation with basis

$$x_1 = (1 \quad 0 \quad 0)^T,\ x_2 = (0 \quad 1 \quad 0)^T,\ x_3 = (0 \quad 0 \quad 1)^T,$$

such that $x_1 \times x_2 = x_{3,}$ $x_1 \times x_3 = -x_{2,}$ $x_2 \times x_3 = x_{1,}$ The only nonzero structure constants are

$$c_{12}^3 = c_{23}^1 = c_{31}^2 = 1, \qquad c_{21}^3 = c_{32}^1 = c_{13}^2 = -1.$$

(The structure constants are unchanged by a cyclic permutation of the indices (123).) The Lie algebra with these structure constants is $\mathfrak{so}(3)$. The Lie group generated by $\mathfrak{so}(3)$ is $SO(3)$, the special orthogonal group. An alternative realization in terms of generators of Lie point symmetries is

$$X_1 = y\partial_x - x\partial_y, \qquad X_3 = \frac{1}{2}(1 + x^2 - y^2)\partial_x + xy\partial_y, \qquad X_3 = xy\partial_x + \frac{1}{2}(-x^2 + y^2)\partial_y.$$

It follows from previous results that structure constants are unaffected by a change of variables or prolongation. HOWEVER, structure constants depend on the choice of basis for $\mathcal{L}$. Life is easier in a basis with as few nonzero structure constants as possible. (This reminds me of Gram-Schmidt orthonormalization process.) If all the basis vectors of the basis commute, the Lie algebra is ***abelian***.

Example 2.11—(A practice problem) Consider the most general two-dimensional Lie algebra with basis $\{X_1, X_2\}$. The commutator of $X_1$with X_2 is of the form $[X_1, X_2] = c_{12}^1 X_1 + c_{12}^2 X_{2.}$ The Lie algebra is abelian iff $c_{12}^1 = c_{12}^2 = 0$. This is also the condition for linear independence: X_1, X_2 are linearly independent if $a = b = 0$ is the only way to make $aX_1 + bX_2 = 0$. If the Lie algebra is not abelian we can find **another** basis $[\check{X}_1, \check{X}_2]$ such that $[\check{X}_1, \check{X}_2] = \check{X}_1$. (We have shown that our algebra is independent of our coordinate basis, so if $[\check{X}_1, \check{X}_2] = \check{X}_1$ then $[X_1, X_2] = X_1$) To compute this **other** basis, note that the commutator of any two generators must be a multiple of $c_{12}^1 X_1 + c_{12}^2 X_2$, the most general commutator by our construction for $X_1$and X_2. Let $\check{X}_1 = c_{12}^1 X_1 + c_{12}^2 X_{2.}$ Suppose $c_{12}^1 \neq 0$ then

$$[\check{X}_1, X_2] = [c_{12}^1 X_1 + c_{12}^2 X_{2,} X_2] = c_{12}^1 [X_1, X_2] + c_{12}^2 [X_2, X_2] = c_{12}^1 [X_1, X_2] = c_{12}^1 [\check{X}_1, \check{X}_2] = c_{12}^1 \check{X}_1$$

which is off by a factor of c_{12}^1. To fix this, let $\check{X}_2 = \frac{1}{c_{12}^1} X_2$. Similarly if $c_{12}^2 \neq 0$ but the Lie algebra is non-abelian, then letting $\check{X}_2 = -\frac{1}{c_{12}^1} X_1$would satisfy our $[\check{X}_1, \check{X}_2] = \check{X}_1$. Let's check.

$$[\check{X}_1, \check{X}_2] = [c_{12}^1 X_1 + c_{12}^2 X_{2,} \check{X}_2] = -\frac{1}{c_{12}^1} [c_{12}^1 X_1 + c_{12}^2 X_{2,} X_1] = -\frac{c_{12}^2}{c_{12}^1} [X_1, X_2] = X_1 \text{ if } c_{12}^1 = 1, c_{12}^2 = -1.$$

Vector spaces can be built from vector subspaces in various ways and Lie algebras may be built from Lie subalgebras. How we join the Lie subalgebras determines the structure constants of our Lie algebra. Let $[\mathcal{M}, \mathcal{N}]$ be the set of all commutators of generators in $\mathcal{M} \subset \mathcal{L}$ with generators $\mathcal{N} \subset \mathcal{L}$, that is, $[\mathcal{M}, \mathcal{N}] = \{[X_i, X_j] : X_i \in \mathcal{M}, X_j \in \mathcal{N}\}$. A subspace $\mathcal{M} \subset \mathcal{L}$ is a subalgebra of a Lie algebra if it's closed under commutation, $[\mathcal{M}, \mathcal{M}] \subset \mathcal{M}$. A subalgebra $\mathcal{M} \subset \mathcal{L}$ is an **ideal** if $[\mathcal{M}, \mathcal{L}] \subset \mathcal{M}$. The set $\{0\}$ and $\mathcal{L}$ are both, trivially, subalgebras and ideals. Any other ideal other than the trivial ideal is a ***proper ideal***. Facts: every one-dimensional subspace of $\mathcal{L}$ is a subalgebra but not necessarily an ideal as each commutator commutes with itself. Almost all Lie algebras of dimension $R \geq 2$ have at least one two-dimensional subalgebra; the one exception (important to physics) is the simple Lie algebra $\mathfrak{so}(3)$.

Example 2.12—(Another Lie algebra important for physics $SL(2)$. Given the three-dimensional Lie algebra $X_1 = \partial_x,\ X_2 = x\partial_x,\ X_3 = x^2\partial_x$. The nontrivial (nonzero) commutators of $[X_i, X_j], i < j$ are

$$[X_1, X_2] = X_1, \qquad [X_1, X_3] = 2X_2, \qquad [X_i, X_j] = X_3.$$

This Lie algebra is the special linear group $\mathfrak{sl}(2)$ which generates $SL(2)$. The subalgebras are Span(X_1, X_2) and Span(X_2, X_3), which are two-dimensional. Span(X_2, X_3) is not closed. Hence it's not a subalgebra. The group $\mathfrak{sl}(2)$ has only the trivial ideals. Note—In later work we use groups (with only one binary operation) to study structure constants versus ideals (with two binary operations).

Important: Given any Lie algebra $\mathcal{L}$, one ideal that can always be constructed is the ***derived subalgebra*** $\mathcal{L}^{(1)}$, which is defined by $\mathcal{L}^{(1)} = [\mathcal{L}, \mathcal{L}]$. Of course $[\mathcal{L}^{(1)}, \mathcal{L}] \subset \mathcal{L}^{(1)}$, making $\mathcal{L}^{(1)}$ an ideal of $\mathcal{L}$. If $\mathcal{L}^{(1)} \neq \mathcal{L}$, we may continue to find the derived subalgebra of $\mathcal{L}^{(1)}$ via $\mathcal{L}^{(2)} = [\mathcal{L}^{(1)}, \mathcal{L}^{(1)}]$, … , $\mathcal{L}^{(k)} = [\mathcal{L}^{(k-1)}, \mathcal{L}^{(k-1)}]$ until we obtain no new subalgebra. If for some k we get $\mathcal{L}^{(k)} = \{0\}$, then $\mathcal{L}$ is said to be ***solvable***. Given R point symmetries of and ODE, an R-dimensional Lie algebra is solvable if there is a chain subalgebras such that $\{0\} = \mathcal{L}_0 \subset \mathcal{L}_1 \subset \cdots \subset \mathcal{L}_R = \mathcal{L}$, where $\dim(\mathcal{L}_k) = k$, such that $\mathcal{L}_{k-1}$ is an ideal of $\mathcal{L}_k$ for each k. Seen from several previous examples we now see any non-abelian two-dimensional Lie algebra has a basis such that $[X_1, X_2] = X_1$, and hence is solvable. For each $R \geq 3$ there exit lie algebras that are not solvable, *e.g.*, $\mathfrak{so}(3)$.

Important: Given an R-dimensional solvable Lie algebra, it's best to choose a basis such that

$$X_k = \mathcal{L}_k, \qquad X_k \neq \mathcal{L}_{k-1}, \qquad k = 1, \dots, R.$$

Hence $\mathcal{L}_k = \mathrm{Span}(X_1, \dots, X_k)$. This basis is the ***canonical basis***. Equivalently, a basis is canonical if $c_{ij}^k = 0\ \ \forall\, i < j \leq k$.

Example 1.13—Recall this example? Given $y^{(iv)} = y'''^{\frac{4}{3}}$, the linearized symmetry condition leads to the following set of infinitesimal generators: $X_1 = \partial y$, $X_2 = x\partial y$, $X_3 = x^2\partial y$, $X_4 = \partial x$, $X_5 = x\partial x$, leading to a five-dimensional ***Lie algebra*** $\mathcal{L} = \{X_1, X_2, X_3, X_4, X_5\}$. From all possible commutators taken from the set $\{X_1, X_2, X_3, X_4, X_5\}$ we only get back the infinitesimal generators $\{X_1, X_2, X_3, X_4\}$. There is no $[X_i, X_j] = \mathrm{c} \cdot X_5$ where c is nonzero. We say that the infinitesimal generators $\{X_1, X_2, X_3, X_4, X_5\}$ form a ***derived subalgebra*** $\mathcal{L}^{(1)} = \mathrm{span}(X_1, X_2, X_3, X_4)$ under the (binary) operation of commutation. Now by taking all possible commutators from the set $\{X_1, X_2, X_3, X_4\}$ you only get back two infinitesimal generators, namely, $\mathcal{L}^{(2)} = \mathrm{span}\{X_1, X_2\}$. Repeating this using all possible commutators from $\{X_1, X_2\}$, you get $\mathcal{L}^{(3)} = \{0\}$. We conclude that $\mathcal{L}$ is solvable.

Finally we get to stepwise integration of ODEs (as to when this is possible at least in principle). We know that given and ODE of order R that has an R-dimensional Lie algebra $\mathcal{L}$, we may rewrite the ODE in terms of differential invariants as

$$v_R = F(r_R)$$

for some function$F$. When can we solve the ODE by using each symmetry generator in turn? (As we proceed, we assume that generators are sufficiently prolonged to describe the linearized group action on all variables.)

Suppose that the generators $X_1, \ldots, X_{R-1}$ form a subalgebra of $\mathcal{L}$. Let (r_{R-1}, v_{R-1}) be the fundamental differential invariants of this subalgebra. If the remaining generator X_R acts on (r_{R-1}, v_{R-1}) as a generator of point transformations, there exist canonical coordinates

$$(r_R, s_R) = \big(r_R(r_{R-1}, v_{R-1}), s_R(r_{R-1}, v_{R-1})\big),$$

as every noninvariant "point" in terms of which $X_R = \partial_{s_R}$. (We can find (r_R, s_R) by methods we've done in the examples.) The v_R is a function of r_R and $s_R = \dot{G}(r_R)$ for some function G. Hence we obtain

$$s_R(r_{R-1}, v_{R-1}) = \int^{r_R(r_{R-1}, v_{R-1})} G(r_R) dr_R + c,$$

which is invariant under the group generated by $X_1, \dots, X_{R-1}$. If this equation can be solved for v_{R-1}, we get a problem of the form $v_{R-1} = F(r_{R-1})$. Provided we can iterate this procedure sufficiently many times, we will obtain the general solution of the ODE. This is so iff $\mathcal{L}$ is solvable.

Proof: Clearly X_R acts on $r_R(r_{R-1}, v_{R-1})$ as a generator of point transformations if the ***restriction*** of X_R to the variables (r_{R-1}, v_{R-1}) is of the form

$$X_R = \alpha(r_{R-1}, v_{R-1})\partial_{r_{R-1}} + \beta(r_{R-1}, v_{R-1})\partial_{v_{R-1}}$$

for some functions α, β, at least one of which is nonzero. Thus we require that

$$X_R r_{R-1} = \alpha(r_{R-1}, v_{R-1}), \qquad X_R v_{R-1}(r_{R-1}, v_{R-1}).$$

The differential invariants r_{R-1}, v_{R-1} satisfy

$$X_i r_{R-1} = 0, \qquad X_R v_{R-1} = 0 \;\; \forall\, i = 1, \dots, R-1,$$

and hence

$$[X_i, X_R] = X_i \alpha(r_{R-1}, v_{R-1}) = 0, \;\forall\, i = 1, \dots, R-1.$$

We may rewrite this as

$$c_{iR}^k X_k r_{R-1} = 0, \qquad \forall\, i = 1, \dots, R-1,$$

leading to

$$c_{iR}^R \alpha(r_{R-1}, v_{R-1}) = 0, \qquad \forall\, i = 1, \dots, R-1,$$

By a similar argument, $[X_i, X_R]v_{R-1} = 0$ leads to

$$c_{iR}^{R}\beta(r_{R-1}, v_{R-1}) = 0, \qquad \forall\, i = 1, \dots, R-1,$$

Therefore, since at least one α, β is nonzero,

$$c_{iR}^{R} = 0, \qquad \forall\, i = 1, \dots, R-1,$$

Span$(X_1, \dots, X_{R-1})$ is a subalgebra if

$$c_{ij}^{R} = 0, \ \ \forall\ 1 \le i < j \le R-1.$$

So $\dots, X_R$ acts as a generator of point transformations on (r_{R-1}, v_{R-1}) iff

$$c_{ij}^{R} = 0, \ \ \forall\ 1 \le i < j \le R.$$

This condition enables us to reduce the order once. Similarly, a second reduction of order is possible if

$$c_{ij}^{R-1} = 0, \ \ \forall\ 1 \le i < j \le R-1.$$

Continuing in this way, each generator X_k may be used to carry out one integration if

$$c_{ij}^{k} = 0, \ \ \forall\ 1 \le i < j \le k.$$

This is only satisfied (in any canonical basis iff $\mathcal{L}$ is solvable. This also works for $R > n$ provided that $\mathcal{L}$ has an n-dimensional solvable algebra (see Hydon text chapter 6 for examples).

IMPORTANT: The following material is on the classification of invariant solutions and the classification of discrete symmetries. The latter material (which you should at least peruse if you're pressing on to working more PDEs with symmetry methods) is of foundational importance to students of particles and fields, which is typically presented to students without any connection to differential equations, and usually by uninformed lecturers and dismal, dry, confusing books. In this section you will with study the classification of invariant solutions and the classification of discrete symmetries in terms of realizations. Several chapters further on will carefully build up the parallel material in terms of matrix representations with, naturally, plenty of step-by-step examples. I don't think that one can be a complete physicist, applied or theoretical, without this much broader picture of symmetry methods. This stuff underlies many of the "invariants" that are just given to you in text books and by lecturers. It also underlies the very mathematics that you will use during your one year studies in graduate quantum physics, not to mention quantum field theories. The applications of these methods, in fact, go to all physics.

On the Classification of Invariant Solutions. Two invariant solutions are equivalent if one can be mapped to the other by a point symmetry of the PDE (the same methods we have developed for ODEs are extended to PDEs in the next chapter which goes back to applications). The set of all invariant solutions of a given differential equation belong to the same equivalence class. Classification of invariant solutions greatly simplifies the search for invariant solutions.

In chapter 1 there was a section dedicated to the characteristic equation with an example applied to the Riccati equation. I used the characteristic equation again in example 1.9. Let's go through one more ODE example, but this time with multiple generators to find invariant solutions. Then let's go through an example for a PDE so that we are fresh when we study the classification of invariant solutions.

Many ODEs can't be completely solved using their Lie point symmetries. We still may be able to derive solutions that are invariant under the group generated by a particular generator X. Recall from chapter 1 that every curve C on the (x, y) plane that is invariant under the group generated by X satisfies $Q(x, y, y') = \eta - y'\xi = 0$ on C. The results that we derive from this equation need to be substituted into the ODE to check if they are indeed solutions (invariant solutions).

Example 2.13—The Blasius equation is $y''' = -yy''$ (so *many* ways to get invariant solutions—ouch.) By application of the linearized symmetry conditions we get the translational symmetry $X_1 = \partial_x$ and the scaling symmetry $X_2 = x\partial_x - y\partial_y$. We may use these symmetries to reduce the order of this ODE to a first order ODE (whose solution is not known). Here we concentrate on finding the invariant solutions using the characteristic equation. Beginning with $X = X_1$, with $\eta = 0$ and $\xi = 1$, the characteristic equation $Q(x, y, y') = \eta - y'\xi = 0$ becomes $Q(x, y, y') = -y' = 0$. The solution to this is $y = c$, where c is an arbitrary constant. Plugging in $y = c$ into the Blasius equation shows that $y = c$ is indeed a solution (an invariant curve actually). Now using the other generator X_2 we have $\eta = -y$ and $\xi = x$. So our characteristic equation is $\eta - y'\xi = -y - xy' = 0$. The solution to this equation is $y = c/x$ (this is the candidate invariant curve). Plugging in our candidate solution into the Blasius equation yields $y''' = -\frac{6c}{x^4} = -yy'' = -y \cdot \frac{2c}{x^3}$. Solving for y gives us $y = 3/x$. This is a solution and invariant curve of the Blasius equation. This was a good review, but we are not quite done. We haven't fully considered all of the possible one-parameter groups. Every remaining one-parameter Lie group is generated by $X = kX_1 + X_2$ (a linear combination of X_1, X_2 where k is a nonzero constant. In this case $\eta = -y$ but $\xi = 1 + k$. (<u>We didn't look at any potentially remaining Lie point symmetries in our examples in chapter 1 or 2</u> (go back and do so if you wish). So now our characteristic equation becomes $Q(x, y, y') = -y - (1 + k)y' = 0$. One solution is $y = 0$. Another solution is $y = \frac{3}{x+k}$. I hope the use of $X = kX_1 + X_2$ helps motivate why we cared so much about the vector space/Lie algebra of the Lie point symmetries of an ODE in our recent sections. Here comes more intuitive meaning of infinitesimal generators which we did cover in chapter 1, but you might have forgotten about. For X_1, we know the tangent vector components are

$$\xi(x, y) = \left(\frac{d\hat{x}}{d\varepsilon}\right)_{\varepsilon=0} = 1 \text{ and } \eta(x, y) = \left(\frac{d\hat{y}}{d\varepsilon}\right)_{\varepsilon=0} = 0, \quad \varepsilon \in \mathbb{R}.$$

By inspection we can readily see that $\hat{x} = x + \varepsilon$ and $\hat{y} = y$. Hence the action of X_1on $y = 3/x$ is to produce $\hat{y} = 3/(\hat{x} - \varepsilon)$, which is invariant under $X_2 = (X_2\hat{x})\partial_{\hat{x}} + (X_2\hat{y})\partial_{\hat{y}} = \left((x\partial_x - \partial_y)(x + \varepsilon)\right)\partial_{\hat{x}} + \left((x\partial_x - y\partial_y)y\right)\partial_{\hat{y}} = x\partial_{\hat{x}} - y\partial_{\hat{y}} = (\hat{x} - \varepsilon)\partial_{\hat{x}} - \hat{y}\partial_{\hat{y}}$. Let us know introduce a very useful notation; if $X_i = \xi_i(x, y)\partial_x + \eta_i(x, y)\partial_y$, then let

$$\hat{X}_i = \xi_i(\hat{x}, \hat{y})\partial_{\hat{x}} + \eta_i(\hat{x}, \hat{y})\partial_{\hat{y}}.$$

Notice then that $\hat{y} = 3/(\hat{x} - \varepsilon)$ is invariant under $X_2 = \hat{X}_2 - \varepsilon\hat{X}_1$. Dropping carets, $y = 3/x$maps to $y = 3/(x - \varepsilon)$, which is invariant under $X_2 - \varepsilon X_1$. A similar calculation shows that $y = 0$ is mapped to itself under X_1.

The invariant canonical coordinate $r(x, y)$satisfies $\xi D_x r + Q r_y = \xi r_x + \eta r_y = 0$, so every invariant solution on which $\xi \neq 0$ is of the form $r(x, y) = c$. There may also be invariant solutions $y = f(x)$ such that $\xi(x, f(x)) = \eta(x, f(x)) = 0$.

Generally speaking, these solutions are gotten by solving either $\xi(x, y) = 0$ or $\eta(x, y) = 0$, then checking (by substitution) that the solutions indeed satisfy the ODE and $\xi(x, f(x)) = \eta(x, f(x)) = 0$. There is yet another way to find invariant solutions on which ξ does not vanish which is useful if the characteristic equation is hard to solve (recall the second example of chapter 1). For Lie point symmetries ξ and η are functions of x and y only. Clearly then the characteristic equation holds if

$$y' = \frac{\eta(x, y)}{\xi(x, y)}$$

on invariant curves for which ξ is not zero. Higher derivatives are calculated by the prolongation formula. (See example 4.7 Hydon text). Can you begin to see why classification of solutions in terms of equivalence relations might be a good idea?

A note on symmetry methods for PDEs. The detailed material for symmetry methods for PDEs lays ahead in a later chapter. Suffice it to say for now that every concept we have developed for ODEs extends to PDEs. Namely, we search for Lie point symmetries via the linearized symmetry condition. With the symmetries in hand we may, as we just did in the previous example, then look for invariant solutions. We may, additionally, try to reduce the order of our PDE, rely on various ansätze, etcetera. As for invariant solutions, we have the same proliferation issue as we do with ODEs.

Equivalence of Invariant Solutions (general discussion). Two invariant solutions are ***equivalent*** if one can be mapped to the other by a point symmetry of the ODE or PDE. Equivalent solutions belong to the same equivalence class. Classification of invariant solutions into equivalence classes greatly simplifies the problem of determining all invariant solutions. To keep things simple, let's restrict ourselves to the problem of the equivalence of solutions that are invariant under a one-parameter Lie group of point symmetries, and to avoid too much clutter, let x and u denote the Nindependent and Mdependent variables respectively, and let z be the set of all variables, $z = (x, u)$. Let

$$\Gamma: z \longrightarrow \hat{z} \tag{38}$$

be a symmetry that acts on a solution that is invariant under the one-parameter group generated by $X = \kappa^i X_i$ where each κ^i is a constant, and the generators $X_i = \zeta_i^S(z)\partial_{z^S}$ form a basis for the Lie algebra. Let a carat over a function indicate that we have changed z by $\hat{z}$. Then, for example, $\hat{X}_i = \zeta_i^S(\hat{z})\partial_{\hat{z}^S}$ is the same function as before, but with the new argument.

Having just worked example 2.13, we should be fresh on the procedure for generating invariant solutions. Suppose $u = f(x)$ is invariant under X. In terms of $(\hat{x}, \hat{u})$ we have $\hat{u} = \tilde{f}(\hat{x})$. The symmetry generator X can also be written in terms of $(\hat{x}, \hat{u})$, then we may remove the carats. In the transformed coordinates, $\hat{u} = \tilde{f}(\hat{x})$ is invariant under the generator $\tilde{X}$. By our construction, u and $\hat{u}$ are equivalent and X and $\tilde{X}$ because a symmetry, Γ maps one to the other respectively. To classify invariant solutions we will classifying the associated symmetry generators. When this is done, one generator from each equivalence class may be used to obtain the invariant solutions. A set containing exactly one generator from each class is an ***optimal system of generators***.

To classify generators we write X in terms of $\hat{z}$. Instead of asking you to recall from chapter 1 that (where F is any smooth function and $\hat{z} = e^{\varepsilon X_j} z$)

$$e^{\varepsilon X_j} F(z) = F\left(e^{\varepsilon X_j} z\right) = F(\hat{z}),$$

I'll just cut and paste the brief material here:

How is the infinitesimal generator affected by a change of coordinates? Suppose (u, v) are new coordinates and let $F(u, v)$ be an arbitrary smooth function. By the chain rule

$$XF(u,v) = XF\big(u(x,y), v(x,y)\big) = \xi(x,y)\partial_x F + \eta(x,y)\partial_y F$$

$$= \xi\left[\frac{\partial F}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial x}\right] + \eta\left[\frac{\partial F}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial F}{\partial v}\frac{\partial v}{\partial y}\right] = \xi[u_x F_u + v_x F_v] + \eta\left[u_y F_u + v_y F_v\right]$$

$$= \left(\xi u_x + \eta u_y\right)F_u + \left(\xi v_x + \eta v_y\right)F_v = (Xu)F_u + (Xv)F_v.$$

Without loss of generality, since $F(u, v)$ is arbitrary, in the new coordinates

$$X = (Xu)\,\partial_u + (Xv)\,\partial_{v.}$$

Thus X represents the tangent vector field in all coordinate systems. If we regard $\{\partial_x, \partial_y\}$ as a *basis* for the space of vector fields on the plane, X is the tangent vector at (x, y). The infinitesimal generator provides a coordinate free way of characterizing the action of Lie symmetries on functions.

If $(u, v) = (r, s)$ are canonical coordinates, the tangent vector is (0,1) and $X = \partial_s$. Let $G\big(r(x,y), s(x,y)\big)$ be a smooth function and $F(x,y) = G\big(r(x,y), s(x,y)\big)$. At any invariant point

(x, y), the Lie symmetries map $F(x, y)$ to $F(\hat{x}, \hat{y}) = G(\hat{r}, \hat{s}) = G(r, s + \varepsilon)$. Applying Taylor's theorem and given $X = \partial_s$, we get

$$F(\hat{x}, \hat{y}) = \sum_{j=0}^{\infty} \frac{\varepsilon^j}{j!} \frac{\partial^j G(r, s)}{\partial s^j} = \sum_{j=0}^{\infty} \frac{\varepsilon^j}{j!} X^j G(r, s).$$

Reverting back to (x, y) coordinates, $F(\hat{x}, \hat{y}) = \sum_{j=0}^{\infty} \frac{\varepsilon^j}{j!} X^j F(x, y)$. If the series converges it is called the Lie series of F about (s, y). We have assumed that (s, y) is not an invariant point, but the expansion is also valid at all invariant points. At an invariant point $X = 0$, and only the $j = 0$ term survives, which is $F(x, y)$. We may express all of this in shorthand to $F(\hat{x}, \hat{y}) = F(e^{\varepsilon X} x, e^{\varepsilon X} y) = e^{\varepsilon X} F(x, y)$.

If $\hat{z} = e^{\varepsilon X_j} z$ and F is any smooth function, we rewrite X in terms of $\hat{z}$.

$$e^{\varepsilon X_j} F(z) = F\left(e^{\varepsilon X_j} z\right) = F(\hat{z}),$$

More generally we define the action of any symmetry $\Gamma: z \longrightarrow \hat{z}$ on any smooth function F similarly

$$\Gamma F(z) = F(\Gamma z) = F(\hat{z}) \tag{39}$$

This will allow us to deal with discrete symmetries as well as Lie point symmetries. Now let $\Gamma(\delta)$ denote the one-parameter Lie group of symmetries generated by X where δ is the group parameter and $\Gamma(\delta): z \longrightarrow e^{\delta X} z$. If F is an arbitrary smooth function, then from (39), $F(z) = \Gamma^{-1} F(\hat{z})$, so multiplying both sides by ΓX produces $\Gamma X F(z) = \Gamma X \Gamma^{-1} F(\hat{z})$, but $XF(z) = F(\hat{z})$, so the equation above becomes

$$\Gamma F(\hat{z}) = \Gamma X F(z) = \Gamma X \Gamma^{-1} F(\hat{z}).$$

By looking at the last term of the equation above us it seems that

$$\hat{X} = \Gamma X \Gamma^{-1}. \tag{40}$$

$$\hat{X} F(\hat{z}) = \Gamma X F(z) = \Gamma X \Gamma^{-1} F(\hat{z}).$$

So what is $\widehat{\mathrm{X}}\widehat{\mathrm{X}}F(\hat{z}) = \widehat{\mathrm{X}}^2 F(\hat{z})$? (This math here is **very** present in quantum physics and field theories.)

$$\widehat{\mathrm{X}}^2 F(\hat{z}) = \Gamma X \Gamma^{-1}\left(\Gamma X \Gamma^{-1} F(\hat{z})\right) = \Gamma X \Gamma^{-1} \Gamma X \Gamma^{-1} F(\hat{z}) = \Gamma X^2 \Gamma^{-1} F(\hat{z}).$$

Continuing on and assuming convergence we can form the Lie series $\hat{\Gamma}(\delta)F(\hat{z}) = e^{\delta\hat{X}}F(\hat{z}) = \Gamma e^{\delta X}\Gamma^{-1}F(\hat{z}) = \Gamma\Gamma(\delta)\Gamma^{-1}F(\hat{z})$. Now since F is arbitrary, it must be that $\Gamma X \Gamma^{-1}$is the generator of the one-parameter Lie group of symmetries $\hat{\Gamma}(\delta) = \Gamma\Gamma(\delta)\Gamma^{-1}$. We will use this result when we get to classifying the discrete symmetries of a differential equation.

(**Now for getting particular**) In this section we ignore generators that depend on arbitrary functions, such as the infinite dimensional subalgebras that occur in linear or linearizable PDEs. To study the equivalence problem we restrict ourselves to Lie symmetries generated by a finite-dimensional Lie algebra with basis $\{X_1, \dots, X_R\}$. In this case it can be shown that the equivalence problem is solvable by studying a finite sequence of one-dimensional problems. In each of these problems, we look at the equivalence under the symmetries obtained from one of the generators in the basis,

$$\Gamma\colon \mathrm{z} \longrightarrow \hat{\mathrm{z}} = e^{\varepsilon X_j} z. \tag{41}$$

From (40)

$$\hat{X} = e^{\varepsilon X_j} X e^{-\varepsilon X_j} \tag{42}$$

for any generator X. (Here, and in the rest of this section we don't sum over the index j as X_j denotes a specific generator.) In particular, (42) holds for $X = X_j$, which commutes with $e^{-\varepsilon X_j}$; thus

$$\hat{X}_j = X_j. \tag{43}$$

We can now write any generator X in terms of $\hat{z}$ by solving (42) for X using (43) to obtain

$$X = e^{-\varepsilon \hat{X}_j} \hat{X} e^{-\varepsilon \hat{X}_j}. \tag{44}$$

The RHS of (44) generates a one-parameter symmetry group under which $\hat{u} = \tilde{f}(\hat{x})$ is invariant under the group generated by

$$\tilde{X} = e^{-\varepsilon X_j} X e^{-\varepsilon X_j}. \qquad (45)$$

which is equivalent to X (under this similarity transformation). This holds for all Lie point symmetry generators X_j. Essentially, the classification problem for generators is solved by using equation (45) with the various generators X_j in turn, to reduce every such generator to its simplest form. How we do this is what comes up next: theory followed by examples.

(**More quantum physics and quantum field theory math**) The equivalence relation (45) involves symmetry generators, rather than any particular solutions of any particular differential equation. Once we've classified the generators for a particular Lie algebra, the classification results apply to all differential equations (ODEs or PDEs) with that Lie algebra. (Here is where a historical divorce happened between Lie algebras and differential equations. You can study these two subjects apart, as we do today, to your great detriment.) Identifying all possible Lie algebras has been solved for scalar ODEs, not for PDEs or systems of ODEs. We usually need to do classification work on a case-by-case basis.

From (45), $\hat{X}$ satisfies the initial value problem (just do it)

$$\left(\frac{d\tilde{X}}{d\varepsilon}\right)_{\varepsilon=0} = \frac{d\tilde{X}}{d\varepsilon}\left(e^{-\varepsilon X_j}\hat{X}e^{-\varepsilon X_j}\right)_{\varepsilon=0} = -X_j\tilde{X} + \tilde{X}X_j = -[X_j, \tilde{X}], \qquad (\tilde{X})_{\varepsilon=0} = X. \qquad (46)$$

Repeating the differentiation gives us

$$\frac{d^2\tilde{X}}{d\varepsilon^2} = -\left[X_j, \frac{d\tilde{X}}{d\varepsilon}\right] = (-1)^2\left[X_j, [X_j, \tilde{X}]\right],$$

and so on. Taylor's theorem then leads to the following series solution (valid for all ε sufficiently near zero):

$$\tilde{X} = X - \varepsilon[X_j, \tilde{X}] + \frac{\varepsilon^2}{2!}\left[X_j, [X_j, \tilde{X}]\right] - \cdots. \tag{47}$$

If X and X_j commute, then (47) reduces to $\tilde{X} = X \ \forall \ \varepsilon$.

For abelian Lie algebras, all generators commute, so no two linearly independent generators are equivalent. The optimal system of generators contains every generator. For non-abelian Lie algebras that occur in quantum physics, the noncommuting generators are not linearly independent—being "entangled"—the underlying time and space being relativistically "entangled", in turn also "entangling" internal symmetries in quantum field theories. ¿Is entanglement evidence of higher dimensions?

For non-abelian Lie algebras, we will use each basis generator X_j to simplify X by eliminating as many of the constants κ^i as possible in $X = \kappa^i X_i$. Note that we have the freedom to multiply X by a nonzero constant λ. (The group generated by λX is the same as the one generated by X.) This multiplication by λ, causing rescaling, will allow us to make simplifications.

<u>Example 2.14</u>—Consider the non-abelian two-dimensional Lie algebra with basis $\{X_1, X_2\}$ such that $[X_1, X_2] = X_1$. This is the Lie algebra $\mathfrak{a}(1)$. Each generator is of the form

$$X = \kappa^1 X_1 + \kappa^2 X_2.$$

Let's find out which generators are equivalent to X under the group generated by X_1 by plugging into the Lie series (47).

$$\tilde{X} = X - \varepsilon[X_1, X] + \frac{\varepsilon^2}{2!}\left[X_1, [X_1, X]\right] - \cdots$$

$$= \kappa^1 X_1 + \kappa^2 X_2 - \varepsilon[X_1, \kappa^1 X_1 + \kappa^2 X_2] + \frac{\varepsilon^2}{2!}\left[X_1, [X_1, \kappa^1 X_1 + \kappa^2 X_2]\right] - \ldots$$

$$= \kappa^1 X_1 + \kappa^2 X_2 - \varepsilon\kappa^2[X_1, X_2] + \frac{\varepsilon^2}{2!}\left[X_1, [X_1, \kappa^2 X_2]\right] - \cdots = \kappa^1 X_1 + \kappa^2 X_2 - \varepsilon\kappa^2 X_1$$

$$= (\kappa^1 - \varepsilon\kappa^2)X_1 + \kappa^2 X_2.$$

The second order term and higher order terms are all zero. If our goal is to simplify (reduce the number of constants κ^i by as much as possible), let's suppose that $\kappa^2 \neq 0$. Then if we choose $\varepsilon = \kappa^1/\kappa^2$, then X is equivalent to $\kappa^2 X_2$. We may rescale κ^2 to one WLOG. The remaining possibility is that $\kappa^2 = 0$. Then $X = \kappa^1 X_1$, which reduces to $X = X_1$ after we rescale κ^1 to one WLOG. So for our Lie algebra $\mathfrak{a}(1)$, the set $\{X_1, X_2\}$ is an optimal system of generators. That is, every solution that is invariant under a one-parameter group generated by $X$is equivalent to a solution that is invariant under the group generated by one of the generators in the optimal system.

We began with by using $X_1$to solve the equivalence problem. What if we had used X_2? Then,

$$\tilde{X} = X - \varepsilon[X_2, X] + \frac{\varepsilon^2}{2!}\big[X_2, [X_2, X]\big] - \cdots$$

$$= \kappa^1 X_1 + \kappa^2 X_2 - \varepsilon[X_2, \kappa^1 X_1 + \kappa^2 X_2] + \frac{\varepsilon^2}{2!}\big[X_2, [X_2, \kappa^1 X_1 + \kappa^2 X_2]\big] - \ldots$$

$$= \kappa^1 X_1 + \kappa^2 X_2 - \varepsilon\kappa^1[X_2, X_1] + \frac{\varepsilon^2}{2!}\big[X_2, [X_2, \kappa^1 X_1]\big] - \cdots$$

$$= \kappa^1 X_1 + \kappa^2 X_2 + \varepsilon\kappa^1 X_1 - \frac{\varepsilon^2}{2!}\kappa^1 X_1 + \cdots = e^{\varepsilon}\kappa^1 X_1 + \kappa^2 X_2.$$

So X_2 acts on the group by rescaling the X_1component. If $\kappa^2 \neq 0$, we rescale X and set $\kappa^2 = 1$ WLOG. If $\kappa^1 \neq 0$ also, let $\varepsilon = -\ln|\kappa^1|$. With this rescaling work done the simplest we can make $\tilde{X}$ is $\tilde{X} = \pm X_1 + X_2$. We can't reach $\tilde{X} = X_2$ without using X_1.

Typically we use most if not all of the one-parameter groups $e^{\varepsilon X_j}$to produce an optimal system of generators. While the calculations are not bad for low dimensional problems, the complexity of the calculations increases rapidly with increasing dimensionality. This is what motivates a move to matrix methods (amenable to manipulation by linear algebra packages). In terms of components we have

$$X = \kappa^i X_i, \tag{48}$$

$$\tilde{X} = \kappa^i \tilde{X}_i = \kappa^i e^{-\varepsilon X_j} X_i e^{-\varepsilon X_j}. \tag{49}$$

Each of the $\tilde{X}_i$ can be written as linear combinations of the X_m. This is efficiently expressed in matrix form by

$$\tilde{X}_i = \left(A(j,\varepsilon)\right)_i^m X_m \tag{50}$$

for some $R \times R$ matrix $A(j,\varepsilon)$. Let $\tilde{\kappa}^m = \kappa^i \left(A(j,\varepsilon)\right)_i^m$. Then

$$\tilde{X} = \tilde{\kappa}^m X_m. \tag{51}$$

Let us introduce the following row vectors

$$\kappa = (\kappa^1, \dots, \kappa^R),$$

$$\tilde{\kappa} = (\tilde{\kappa}^1, \dots, \tilde{\kappa}^R).$$

Then Γ may be regarded as a mapping that acts on the constants κ as $\Gamma\colon \kappa \longrightarrow \tilde{\kappa} = \kappa A(j,\varepsilon)$. Since $\tilde{X}_i = e^{-\varepsilon X_j} X_i e^{-\varepsilon X_j}$ we have $\frac{d\tilde{X}_i}{d\varepsilon} = -e^{-\varepsilon X_j}\left[X_j, X_i\right] e^{\varepsilon X_j} = c_{ij}^k \left(\tilde{X}_i\right)_{\varepsilon=0} = X_i$. Then from equation (50)

$$\frac{d\left(A(j,\varepsilon)\right)_i^m}{d\varepsilon} X_m = c_{ij}^k \left(A(j,\varepsilon)\right)_i^m X_m, \qquad \left(A(j,0)\right)_i^m X_m = X_i. \tag{52}$$

Let's rewrite the structure constants themselves in terms of a matrix

$$\left(C(j)\right)_i^k = c_{ij}^k.$$

Then,

$$\frac{dA(j,\varepsilon)}{d\varepsilon} X_m = C(j) A(j,\varepsilon). \;\; A(j,0) = I, \tag{53}$$

where I is the identity matrix. The general solution of (53) is

$$A(j,\varepsilon) = e^{\varepsilon C(j)} = \sum_{n=0}^{\infty} C(j)^n \frac{\varepsilon^n}{n!}. \tag{54}$$

Sometimes you may solve (54) by hand, but sometimes you may need a computer algebra system.

As we know, the generators of abelian Lie algebras cannot be simplified at all because each generator is invariant under the group generated by any other one $[X_i, X_j] = 0$. Some non-abelian Lie algebras may also have one or more invariants, $I(\kappa)$, where κ is the vector from $X = \kappa^i X_i$ such that

$$I\left(\kappa e^{\varepsilon C(j)} = \tilde{\kappa}\right) = I(\kappa A(j,\varepsilon) = \tilde{\kappa}) = I(\kappa), \qquad \forall\, k, \varepsilon. \tag{55}$$

¿Huh? In English please! If I place a meter stick in an orthogonal x, y, z-coordinate system its length $\Delta x^2 + \Delta y^2 + \Delta z^2 = 1m$. In any other x', y', z'-coordinate system gotten to via some general matrix transformation that preserves length, it better be that $\Delta x'^2 + \Delta y'^2 + \Delta z'^2 = 1m$. Clearly this places restrictions on the matrices which transform (x, y, z) to (x', y', z'). In this example we require our transformation matrix to have determinant +1 for example.

Mathematically, these invariants $I(\kappa)$ act as constraints on the amount of simplification that is possible, so it is important to be able to derive them systematically. To the physicist, these invariants are loaded with physical meaning, *e.g.*, eigenvalues to differential equations. Differentiating (55) wrt ε at $\varepsilon = 0$ results in both the necessary and sufficient condition for $I(\kappa)$ to be invariant:

$$\kappa C(j) \nabla I(\kappa) = 0, \qquad \forall\, j, \tag{56}$$

where (the matrix PDE) is given by

$$\nabla I(\kappa) = \begin{bmatrix} I_1(\kappa) \\ \vdots \\ I_R(\kappa) \end{bmatrix}, \qquad I_i(\kappa) \equiv \frac{\partial \mathrm{I}(\kappa)}{\partial \kappa^{\mathrm{i}}}.$$

The invariance conditions (55) can be solved using the method of characteristics. Here comes a little more linear algebra. We may rewrite (56) as

$$K(\kappa)\nabla I(\kappa) = \mathbf{0}, \qquad (57)$$

where $K(\kappa)$ is the $R \times R$ matrix whose j^{th} row is $\kappa C(j)$. The matrix PDE (57) can be simplified by reducing $K(\kappa)$ to echelon form; the resulting equations sometimes being readily solvable by the method of characteristics. If $\rho = \mathrm{Rank}\big(K(\kappa)\big)$, there are $R - r$ functionally independent invariants.

Example 2.15—Let's use the matrix method to determine an optimal system of generators for the three-dimensional Lie algebra $\mathfrak{sl}(2)$. Say we start off with the basis $\{X_1, X_2, X_3\}$ such that

$$[X_1, X_2] = X_1, \qquad [X_1, X_3] = 2X_2, \qquad [X_2, X_3] = X_3.$$

What is our goal? This example will have lots of algebraic manipulations. Let's first look at our goal. If you recall our examples, we've used the linearized symmetry condition on differential equations to extract their infinitesimal generators $\{X_1, X_2, X_3\}$, which form a Lie algebra. With these generators we proceeded to crank out invariant solutions. Unless $\{X_1, X_2, X_3\}$ is an optimal system of generators, some invariant solutions may turn out to be invariant to a messy, as yet unknown combination of the X_i. We have seen this in an example 2.13 with $X = kX_1 + X_2$. To avoid this, our goal is to transform $\{X_1, X_2, X_3\}$ to an optimal system of generators $\{\tilde{X}_1, \tilde{X}_2, \tilde{X}_3\}$ so that every solution that is invariant under a one-parameter group generated by $X = \tilde{\kappa}^i \tilde{X}_i$ is equivalent to precisely one of the $\tilde{X}_i$ of the optimal system. In net then, all of the solutions fall into equivalence classes, one equivalence class for each $\tilde{X}_i$. Having done this, one generator from each equivalence class $[\tilde{X}_i]$ is used to obtain the desired set of invariant solutions. When the dust has settled, this example will be very enlightening to physicist and mathematician alike.

Right now we sit with $X = \kappa^1 X_1 + \kappa^2 X_2 + \kappa^3 X_3$. We are looking for the optimal version $\tilde{X} = \tilde{\kappa}^1 \tilde{X}_1 + \tilde{\kappa}^2 \check{X}_2 + \tilde{\kappa}^3 \tilde{X}_3$ where $\tilde{X}$ is a linear combination of the old basis generators $\tilde{X} = \tilde{\kappa}^m X_m$ with the $\tilde{\kappa}^m$ being as simple as possible. Let's label the "old" structure constants by using $\left[X_i, X_j\right] = c_{ij}^k X_k$. In the first case we have $c_{12}^1 = 1$, $c_{21}^1 = -1$. Then we have $c_{13}^2 = 2$, $c_{31}^2 = -2$. In the third case we have $c_{23}^3 = 1$, $c_{32}^3 = -1$. For $j = 1$ we have $c_{21}^1 = -1$ and $c_{31}^2 = -2$. WARNING: Since j is fixed, evidently the elements of the matrix $\left(C(j=1)\right)_i^k$ depend not, as usual on i, j, but on i, k. Look at the first part of equation (52), $\frac{d\left(A(j,\varepsilon)\right)_i^m}{d\varepsilon} X_m = c_{ij}^k \left(A(j,\varepsilon)\right)_i^m X_m$ to convince yourself. So plugging into $\left(C(j=1)\right)_i^k = c_{ij=1}^k$ we get $\left(C(j=1)\right)_{i=2}^{k=1} = -1$ and $\left(C(j=1)\right)_{i=3}^{k=2} = -2$

$$C(j=1) = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -2 & 0 \end{bmatrix}.$$

For $j = 2$ we have $c_{12}^1 = 1$ and $c_{32}^3 = -1$. We get $\left(C(j=2)\right)_{i=1}^{k=1} = 1$ and $\left(C(j=2)\right)_{i=3}^{k=3} = -1$.

$$C(j=2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

For $j = 3$ we have $c_{13}^2 = 2$ and $c_{23}^3 = 1$. We get $\left(C(j=3)\right)_{i=1}^{k=2} = 2$ and $\left(C(j=3)\right)_{i=2}^{k=3} = 1$.

$$C(j=3) = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

According to (54) $A(j,\varepsilon) = e^{\varepsilon C(j)} = \sum_{n=0}^{\infty} C(j)^n \frac{\varepsilon^n}{n!}$. Starting with $j = 1$, what is $e^{\varepsilon C(1)}$?

$$e^{\varepsilon C(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{\varepsilon}{1!}\begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -2 & 0 \end{bmatrix} + \frac{\varepsilon^2}{2!}\begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -2 & 0 \end{bmatrix}^2 + \frac{\varepsilon^3}{3!}\begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -2 & 0 \end{bmatrix}^3 + \cdots$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{\varepsilon}{1!}\begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -2 & 0 \end{bmatrix} + \frac{\varepsilon^2}{2!}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} + \frac{\varepsilon^3}{3!}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The matrix Taylor series terms become the zero matrix for all terms cubic and higher. Thus

$$A(1,\varepsilon)=\begin{bmatrix}1 & 0 & 0\\ -\varepsilon & 1 & 0\\ \varepsilon^2 & -2\varepsilon & 1\end{bmatrix}.$$

Similarly,

$$A(2,\varepsilon)=\begin{bmatrix}e^{\varepsilon} & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & e^{-\varepsilon}\end{bmatrix},\qquad A(3,\varepsilon)=\begin{bmatrix}1 & 2\varepsilon & \varepsilon^2\\ 0 & 1 & \varepsilon\\ 0 & 0 & 1\end{bmatrix}.$$

Let's look for row vectors $\kappa C(j)$ to check for the existence of invariants. Let' start with $\kappa C(1)$

$$[\kappa^1 \quad \kappa^2 \quad \kappa^3]\cdot\begin{bmatrix}0 & 0 & 0\\ -1 & 0 & 0\\ 0 & -2 & 0\end{bmatrix}=[-\kappa^2 \quad -2\kappa^3 \quad 0].$$

For $j=2$,

$$[\kappa^1 \quad \kappa^2 \quad \kappa^3]\cdot\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & -1\end{bmatrix}=[\kappa^1 \quad 0 \quad -\kappa^3].$$

For $j=3$,

$$[\kappa^1 \quad \kappa^2 \quad \kappa^3]\cdot\begin{bmatrix}0 & 2 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix}=[0 \quad 2\kappa^1 \quad \kappa^2].$$

So any invariants satisfy

$$\begin{bmatrix}-\kappa^2 & -2\kappa^3 & 0\\ \kappa^1 & 0 & -\kappa^3\\ 0 & 2\kappa^1 & \kappa^2\end{bmatrix}\cdot\begin{bmatrix}I_1(\kappa)\\ I_2(\kappa)\\ I_3(\kappa)\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}.$$

Note that $\rho=2$, so there is only $R-r=3-2=1$ invariant. Recall from linear algebra the eigenvalue problem $\mathbf{A}x=\lambda x$ is solved by finding the roots of the characteristic equation

$$\mathrm{Det}\begin{bmatrix}a_{11}-\lambda & \cdots & a_{1n}\\ \vdots & \ddots & \vdots\\ a_{n1} & \cdots & a_{nn}-\lambda\end{bmatrix}=0.$$

This means finding the roots of Det $\begin{bmatrix} -\kappa^2-\lambda & -2\kappa^3 & 0 \\ \kappa^1 & -\lambda & -\kappa^3 \\ 0 & 2\kappa^1 & \kappa^2-\lambda \end{bmatrix} = 0$, or $\lambda^2 - (\kappa^2)^2 + 4\kappa^1\kappa^3 = 0$. So our only invariant for $\mathfrak{sl}(2)$ is $I = (\kappa^2)^2 - 4\kappa^1\kappa^3$. (If you're rusty, look this up in a sophomore linear algebra text.) So what the hell is $I = (\kappa^2)^2 - 4\kappa^1\kappa^3$? It is our invariant "meter stick" in our "coordinate system" $\kappa^1, \kappa^2, \kappa^3$. The only way we can affect I is by rescaling X, which is equivalent to multiplying κ by a nonzero constant. Since I is quadratic in the components of κ, rescaling can only multiply I by a positive constant. Thus we must consider three distinct problems: $I > 0$, $I < 0$, $I = 0$.

The vector κ is transformed by the matrices $A(j, \varepsilon)$ as follows:

$$[\kappa^1 \quad \kappa^2 \quad \kappa^3] \cdot \begin{bmatrix} 1 & 0 & 0 \\ -\varepsilon & 1 & 0 \\ \varepsilon^2 & -2\varepsilon & 1 \end{bmatrix} = [\kappa^1 - \varepsilon\kappa^2 + \varepsilon^2\kappa^3 \quad \kappa^2 - 2\varepsilon\kappa^3 \quad \kappa^3], \qquad \text{for } A(1, \varepsilon).$$

What the hell is $[\kappa^1 - \varepsilon\kappa^2 + \varepsilon^2\kappa^3 \quad \kappa^2 - 2\varepsilon\kappa^3 \quad \kappa^3]$? It is a new, transformed coordinate system with

$$\tilde{\kappa}^1 = \kappa^1 - \varepsilon\kappa^2 + \varepsilon^2\kappa^3, \qquad \tilde{\kappa}^2 = \kappa^2 - 2\varepsilon\kappa^3, \qquad \tilde{\kappa}^3 = \kappa^3, \qquad \text{(A)}$$

and in this new coordinate system $I_A = (\tilde{\kappa}^2)^2 - 4\tilde{\kappa}^1\tilde{\kappa}^3$, and $I = I_A$. No way dudes. Let's check it out.

$$(\kappa^2)^2 - 4\kappa^1\kappa^3 = (\kappa^2 - 2\varepsilon\kappa^3)^2 - 4(\kappa^1 - \varepsilon\kappa^2 + \varepsilon^2\kappa^3)\kappa^3.$$

If you do the algebra, indeed $I = I_A$, a meter length in one coordinate system is a meter length in another coordinate system. Let's press to $A(2, \varepsilon)$.

$$[\kappa^1 \quad \kappa^2 \quad \kappa^3] \cdot \begin{bmatrix} e^{\varepsilon} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-\varepsilon} \end{bmatrix} = [e^{\varepsilon}\kappa^1 \quad \kappa^2 \quad e^{-\varepsilon}\kappa^3], \qquad \text{for } A(2, \varepsilon).$$

This is another a new, transformed coordinate system with

$$\tilde{\kappa}^1 = e^{\varepsilon}\kappa^1, \qquad \tilde{\kappa}^2 = \kappa^2, \qquad \tilde{\kappa}^3 = e^{-\varepsilon}\kappa^3, \qquad \text{(B)}$$

and in this new coordinate system $I_B = (\tilde{\kappa}^2)^2 - 4\tilde{\kappa}^1\tilde{\kappa}^3 = I$. Very cool. Let's press to $A(3, \varepsilon)$.

$$[\kappa^1 \quad \kappa^2 \quad \kappa^3] \cdot \begin{bmatrix} 1 & 2\varepsilon & \varepsilon^2 \\ 0 & 1 & \varepsilon \\ 0 & 0 & 1 \end{bmatrix} = [\kappa^1 \quad 2\varepsilon\kappa^1 + \kappa^2 \quad \varepsilon^2\kappa^1 + \varepsilon\kappa^2 + \kappa^3], \qquad \text{for } A(3,\varepsilon).$$

This is our third new, transformed coordinate system with

$$\tilde{\kappa}^1 = \kappa^1, \qquad \tilde{\kappa}^2 = 2\varepsilon\kappa^1 + \kappa^2, \qquad \tilde{\kappa}^3 = \varepsilon^2\kappa^1 + \varepsilon\kappa^2 + \kappa^3, \qquad \text{(C)}$$

and in this new coordinate system $I_C = (\tilde{\kappa}^2)^2 - 4\tilde{\kappa}^1\tilde{\kappa}^3 = I$. Very cool yet again.

Now our task is to use the fact that in the original coordinate system, (A), (B), and (C) we have

$$I = (\kappa^2)^2 - 4\kappa^1\kappa^3 = I_A = (\kappa^2 - 2\varepsilon\kappa^3)^2 - 4(\kappa^1 - \varepsilon\kappa^2 + \varepsilon^2\kappa^3)\kappa^3 = I_B = (\kappa^2)^2 - 4e^{\varepsilon}\kappa^1 \cdot e^{-\varepsilon}\kappa^3$$

$$= I_C = (2\varepsilon\kappa^1 + \kappa^2)^2 - 4\kappa^1 \cdot (\varepsilon^2\kappa^1 + \varepsilon\kappa^2 + \kappa^3)$$

to simply out an optimal systems of generators.

Now suppose $I > 0$. The first and third components of vector (A) involve the parameter ε in addition to the set $\{\kappa^1, \kappa^2, \kappa^3\}$. We are looking to find the simplest form of vector (A) and we have the freedom to play with ε. Suppose $\kappa^3 \neq 0$ and look at the first component of vector (A), $\kappa^1 - \varepsilon\kappa^2 + \varepsilon^2\kappa^3$. What does ε have to be so that the first component of vector (A) is zero. To ask this is to solve for ε in the quadratic equation $\varepsilon^2\kappa^3 - \varepsilon\kappa^2 + \kappa^1 = 0$. Rearranging leads to

$$\varepsilon^2 - \frac{\kappa^2}{\kappa^3}\varepsilon + \frac{\kappa_1}{\kappa_3} = 0$$

The quadratic formula leads to

$$\varepsilon = \frac{\kappa^2}{2\kappa^3} \pm \sqrt{\left(\frac{\kappa^2}{\kappa^3}\right)^2 - 4\frac{\kappa_1}{\kappa_3}} = \frac{\kappa^2}{2\kappa^3} \pm \frac{1}{2\kappa^3}\sqrt{(\kappa^2)^2 - 4\kappa^1\kappa^3} = \frac{\kappa^2}{2\kappa^3} \pm \frac{\sqrt{I}}{2\kappa^3}.$$

Pick the positive root to let

$$\varepsilon = \frac{\kappa^2 + \sqrt{I}}{2\kappa^3}.$$

So now with ε as selected, vector (A) is reduced to

$$\tilde{\kappa}^1 = 0, \qquad \tilde{\kappa}^2 = \kappa^2 - 2\varepsilon\kappa^3, \qquad \tilde{\kappa}^3 = \kappa^3.$$

If instead it happened that $\kappa^3 = 0$ without saying anything about κ^1, then let $\varepsilon = \kappa^1/\kappa^2$. This leaves us vector (A) looking like if $\kappa^3 \neq 0$:

$$\tilde{\kappa}^1 = 0, \qquad \tilde{\kappa}^2 = \kappa^2, \qquad \tilde{\kappa}^3 = 0.$$

We may rescale $\kappa^2 = 1$ to get vector $(0,1,0)$. So from equation (50), $\tilde{X}_i = \left(A(j,\varepsilon)\right)_i^m X_m$,

one of our optimal system of generators is $\tilde{X}_2 = X_3$. We have two more to go.

Let us suppose $I < 0$. In our original coordinate system this would mean that $\kappa^1\kappa^3 > 0$. Let's study vector (A). Let $\varepsilon = \kappa^2/2\,\kappa^3$. Then vector (A) reduces to

$$\tilde{\kappa}^1 = \kappa^1 - \varepsilon\kappa^2 + \varepsilon^2\kappa^3, \qquad \tilde{\kappa}^2 = 0, \qquad \tilde{\kappa}^3 = \kappa^3.$$

The κ^2 term has been disappeared. Then in vector (B) let $\varepsilon = \frac{1}{2}\ln\frac{\kappa^3}{\kappa^1}$. Then vector (B) becomes

$$\tilde{\kappa}^1 = \frac{\kappa^1}{(\kappa^3)^2}, \qquad \tilde{\kappa}^2 = 0, \qquad \tilde{\kappa}^3 = \frac{\kappa^1}{(\kappa^3)^2}.$$

WLOG let $\kappa^1 = \kappa^3 = 1$. Then vector (B) becomes $(1,0,1)$. So from equation (50), $\tilde{X}_i = \left(A(j,\varepsilon)\right)_i^m X_m$,

one of our optimal system of generators is $X_1 + X_3$. One more to go.

If $I = 0$ then either all three components of the vector κ are zero, or κ^2 and either of κ^1, κ^3 are zero. Now a property of the matrices which go with the invariant I is that

$$[\kappa^1 \quad \kappa^2 \quad \kappa^3] \cdot \prod A(i,\varepsilon),$$

retains the invariant I, where $\prod A(i,\varepsilon)$ is any product of $A(i,\varepsilon)$ with ε any constant. Suppose $\kappa^1,\ \kappa^2 = 0$. Then vector (C) becomes

$$\tilde{\kappa}^1 = 0, \qquad \tilde{\kappa}^2 = 0, \qquad \tilde{\kappa}^3 = \kappa^3.$$

One of our optimal generators already includes the third component. Now

$$(0 \quad 0 \quad \kappa^3)\, A(1,1)A(3,1) = (0 \quad 0 \quad \kappa^3)\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = (\kappa^3 \quad -2\kappa^3 \quad \kappa^3)\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= (\kappa^3 \quad 0 \quad 0),$$

so we can replace the second and third components of any κ with zeros. We may rescale $\kappa^3 = 1$ to get $(1 \quad 0 \quad 0)$. Thus our last optimal generator is X_1.

Our optimal set of generators are $\{X_1, X_2, X_1 + X_3\}$ for $\mathfrak{sl}(2)$. The optimal set of generators contains every generator. Each generator of the optimal set may now be used to obtain the desired set of invariant solutions for any ODE or PDE with Lie algebra $\mathfrak{sl}(2)$. Don't get the wrong impression from the previous two examples where the number of generators coincides with the dimension R of the Lie algebra. Quite commonly the number of inequivalent generators exceeds R.

We have looked at building an optimal set of generators quite divorced from differential equations. We can proceed to calculate the associated invariant solutions of an ODE or PDE with a given optimal set of generators to derive an ***optimal system of invariant solutions***, but we potentially face two obstacles. First of all the generator in the optimal system might not yield any invariant solutions.

Example 2.16—Consider $y'' = y^{-3}$. Application of the linearized symmetry condition would lead to the following Lie point symmetries

$$X_1 = \partial_x, \qquad X_2 = x\partial_x + \frac{1}{2}y\partial_y, \qquad X_3 = x^2\partial_x + xy\partial_y.$$

If you repeat the work of example 2.15 you will find that these infinitesimal generators have the Lie algebra $\mathfrak{sl}(2)$. The optimal system of generators is $\{X_1, X_2, X_1 + X_3\}$. Alas no solution is invariant under the group generated by X_1. From X_1 we see that $\xi = 1$ and $\eta = 0$, so $dy/dx = 0$, so $y = c$, where c is an arbitrary constant. We need to check if $y = c$ is a solution to our ODE. Nope. The 2nd derivative of a constant is zero. The same applies to the group generated by X_2. In this case $\xi = x$ and $\eta = \frac{1}{2}y$, so $dy/dx = y/2x$, so $y = c\sqrt{x}$. We need to substitute this result into our ODE to see if it is a solution.

$$\text{LHS: } y'' = -\frac{1}{4}\cdot\frac{c}{x^{-\frac{3}{2}}} = \text{RHS} = \frac{c^{-3}}{x^{-\frac{3}{2}}}.$$

The only way to make this work if it happens to be that $c^4 = -4$. This means c is not real. So there are no real-valued solutions for X_2. For $X_1 + X_3$ we have $\xi = 1 + x^2$ and $\eta = \frac{1}{2}y$. Then

$$\frac{dy}{dx} = \frac{1}{2}\cdot\frac{y}{1 + x^2}.$$

So $y = \pm\sqrt{1 + x^2}$. Again, we check to see if this is a solution to our ODE by substituting it in. (Review the Lie series (47) and example 2.13.) The action of the other two symmetries $X_{1,}\, X_2$ yields the general solution of the ODE

$$y = \pm\sqrt{c_1 + (x + c_2)^2/c_1}, \;\; c_1 = 0.$$

Recall that $X_{1,}\, X_2$ are representatives of the classes of generators with $I = 0$ and $I > 0$, respectively. Might there be invariant solutions associated with other generators in these classes? No. None. Zilch.

The second obstacle we might face in calculating the associated invariant solutions to a Lie algebra is that the reduced equation(s) determining one or more invariant solutions may be too difficult to solve analytically. Even if we cannot obtain an optimal system of solutions, we may be able to find some solutions. In the Hydon text there is an example to this case (example 10.4) dealing with a linear

PDE. I have not yet covered symmetry methods for PDEs. Please consult the Hydon book if you wish. We will soon return to working problems dealing with both ODEs and PDEs. Only one more theory section remains in Part I—yay!

Discrete symmetries. (Motivation—taken from Hydon's text.)

One. (Physicist pay attention.) Many nonlinear boundary value problems (BVPs) have multiple solutions, and it is necessary to identify when and how the physical system changes its behavior as any parameters vary. It is important to identify all of the symmetries in a problem in order to understand its behavior fully and correctly. I learned this big time in dealing with thermodynamical systems.

Two. Discrete point symmetries may be used to increase the efficiency of numerical methods. If a BVP is symmetric and the solution is known to be unique (which is quite often the case in physics) then computation can be carried out on a reduced domain. Look up spectral methods which may be used with basis functions that are invariant under the symmetry.

Three. As with Lie symmetries, discrete symmetries may be used to generate new solutions from known solutions. Discrete symmetries may also be used to simplify an optimal system of generators. If two generators are related by a discrete symmetry, you need only one of them.

Four. (Physicist pay attention.) Discrete symmetries involving charge conjugation, parity change, and time reversal (CPT symmetries) are central in quantum field theories. Of course there are many more discrete symmetries underpinning general physics.

Five. (Wow!) We've learned about Lie point symmetries and we're about to learn about discrete point symmetries, but there are many more in physics and in mathematics, such as Legendre contact transformations (such transformations can occur as symmetries of differential equations, even if there are no Lie contact symmetries). Noether's (first) theorem states that any differentiable symmetry of the action of a physical system has a corresponding conservation law. Auto-Buckland transformations (what are these?) are nonpoint discrete symmetries which enable us to construct hierarchies of solutions to nonlinear integrable PDEs. For differential equations without Lie point symmetries, there are C^∞ symmetries. Symmetries are at the heart of phenomenological physics and fundamental physics. Sometimes it is the lab (nature) that leads us to some new symmetry, but it can go the other way around. I begin to feel small, like Isaac Newton at the beach picking up a shiny seashell before a massive ocean of mystery. At times I feel more like a swimmer who has swum off too far and is now being dragged out to open ocean as a tempest begins to brew. I don't have enough of a lifetime to master all the symmetries we know of. They are already far too numerous to keep from slipping through my fingers. The meaning and magnitude of symmetries yet to be pulled from the depths boggle my mind. I suppose this will be so for any being no matter how advanced its civilization, how big and powerful its mind might be, infinity is bigger. Everything that is seems to be the result of fundamental symmetries and subsequent emergent symmetries.

How to obtain discrete symmetries from Lie point symmetries. We begin as before, restricting ourselves to the problem of the equivalence of solutions that are invariant under a one-parameter Lie group of point symmetries. Let x and u denote the N independent and Mdependent variables respectively, and let z be the set of all variables, $z = (x, u)$. Let

$$\Gamma: z \longrightarrow \hat{z} \tag{38}$$

be a symmetry of a differential equation, the Lie algebra $\mathcal{L}$ is R-dimensional, and the generators

$$X_i = \zeta_i^S(z)\partial_{z^S} \tag{58}$$

form a basis where each κ^i is a constant. We showed that if $X \in \mathcal{L}$ then

$$\hat{X} = \Gamma X \Gamma^{-1}. \tag{59}$$

generates a one-parameter Lie group of point symmetries of the differential equation, the Lie algebra $\mathcal{L}$ being the set of all generators of Lie point symmetries. Consequently $\hat{X} \in \mathcal{L}$. In particular, each basis generator

$$\hat{X}_i = \Gamma X_i \Gamma^{-1} = \zeta_i^S(\hat{z})\partial_{\hat{z}^S} \tag{60}$$

is in $\mathcal{L}$. The set of generators $\{\hat{X}_1, \dots, \hat{X}_R\}$ is also a basis for $\mathcal{L}$, but with $\hat{z}$ replacing z. Therefore each X_i may be written as a linear combination of the $\hat{X}_i$'s as follows:

$$X_i = b_i^l \hat{X}_l, \tag{61}$$

the coefficients b_i^l being constants determined by the symmetry Γ and the basis $\{\hat{X}_1, \dots, \hat{X}_R\}$. If there is a diffeomorphism between the z coordinate system and the $\hat{z}$ coordinate system, then the $\frac{\partial \hat{z}^q}{\partial z^r}$ are the $R \times R$ linear terms of this diffeomorphism. So if follows that the $b_i^l = \frac{\partial \hat{z}^l}{\partial z^i}$, $1 \le i, l \le R$. It will help simplify things to regard these b_i^l coefficients as an $R \times R$ matrix $\boldsymbol{B} = (b_i^l)$. Since the linear equations (61) are a transformation between basis of X_i's, matrix B is nonsingular. There are $M + N$ total variables z. Multiply the RHS of equation (60) by $\hat{z}^S$ to get $\hat{X}_i \hat{z}^S = \zeta_i^S(\hat{z})$. Multiply this result on the LHS by b_i^l to get

$$b_i^l \hat{X}_i \hat{z}^S = b_i^l \zeta_i^S(\hat{z}), \qquad 1 \le i, r \le R, \qquad 1 \le s \le M + N.$$

Since $i, l \le R$, the term $b_i^l \zeta_i^S(\hat{z})$ is a linear mapping of the $M + N$ $\zeta_i^S(\hat{z})$'s to the R $\zeta_i^r(z)$'s. We may thus write

$$\zeta_i^r(z)\frac{\partial \hat{z}^s}{\partial z^r} = b_i^l \hat{X}_i \hat{z}^S = b_i^l \zeta_i^S(\hat{z}), \qquad 1 \le i, r \le R, \qquad 1 \le s \le M + N. \tag{62}$$

This system of $(M + N)R$ partial differential equations can be solved the method of characteristics, which yields $\hat{z}$ in terms of z, the unknown constants b_i^l, and some arbitrary constants or functions of integration. This (equation (62)) is so for every symmetry, discrete or continuous. Bear in mind that (62) may have solutions which are not symmetries. How do we check if a solution is not a symmetry? Plug it into (43) of course. As we already know the Lie point symmetries, we may factor them out (remove them) at any convenient stage of the calculation. What results is the list of inequivalent discrete point symmetries that can't be mapped to one another by any Lie point symmetry.

<u>Example 2.17</u>—Consider the ODE $y'' = \tan y'$. We see that With $z = (x, y)$ in this example. From the linearized symmetry condition, the Lie generators are $X_1 = \partial_x,\ X_2 = \partial_y$. Given this, we see from equation (58), $X_i = \zeta_i^S(z)\partial_{z^S}$, that

$$X_1 = \zeta_1^1(x,y)\partial_{z^1=x} + \zeta_1^2(x,y)\partial_{z^2=y} = \partial_x.$$

Then $\zeta_1^1(x,y) = 1$ and $\zeta_1^2(x,y) = 0$. From the second generator we see that

$$X_2 = \zeta_2^1(x,y)\partial_{z^1=x} + \zeta_2^2(x,y)\partial_{z^2=y} = \partial_y.$$

So then $\zeta_2^1(x,y) = 0$ and $\zeta_2^2(x,y) = 1$. From equation (61), $X_i = b_i^l \hat{X}_l$ we have

$$X_1 = b_1^1\hat{X}_1 + b_1^2\hat{X}_1, \qquad X_2 = b_2^1\hat{X}_1 + b_2^2\hat{X}_1.$$

So equation (62) becomes

$$\begin{bmatrix} \hat{x}_x & \hat{y}_x \\ \hat{x}_y & \hat{y}_y \end{bmatrix} = \begin{bmatrix} b_1^1 & b_1^2 \\ b_2^1 & b_2^2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

So we have $\hat{x}_x = b_1^1,\ \hat{x}_y = b_2^1,\ \hat{y}_x = b_1^2,\ \hat{y}_y = b_2^2$. So the most general solution to these are $\hat{x}(x,y) = b_1^1 x + b_2^1 y + c_1,\ \hat{y}(x,y) = b_1^2 x + b_2^2 y + c_2$.

Let's pause and take stock. Here, where we are interested in discovering the *discrete* symmetries. Our emphasis has shifted to the actual relationship between the coordinates in the primed and unprimed systems. When were interested finding invariant solutions, we were interested in finding the invariant coordinates beginning with r. For the first generator, X_1in this example we have,

$\xi(x,y) = \left(\frac{d\hat{x}}{d\varepsilon}\right)_{\varepsilon=0} = 1$ and $\eta(x,y) = \left(\frac{d\hat{y}}{d\varepsilon}\right)_{\varepsilon=0} = 0$. We didn't care about finding the actual symmetry connecting (x,y) to $(\hat{x},\hat{y})$ so long as we had the tangent vector (ξ,η), but it's simple enough to see from inspection that $\hat{x} = x + \varepsilon$. Repeating all of this for X_2 we see that $\xi(x,y) = \left(\frac{d\hat{x}}{d\varepsilon}\right)_{\varepsilon=0} = 0$ and $\eta(x,y) = \left(\frac{d\hat{y}}{d\varepsilon}\right)_{\varepsilon=0} = 1$. By inspection $\hat{y} = y + \varepsilon$ is the simplest solution for $\eta(x,y)$. So we have

$$\hat{x} = x + \varepsilon, \qquad \hat{y} = y + \varepsilon,$$

where ε which might as well be c_1. This is certainly not the most general mapping from (x,y) to $(\hat{x},\hat{y})$. This is what equation (63) gives us. As you can see from $\hat{x} = x + \varepsilon,\ \hat{y} = y + \varepsilon$, the generators X_1, X_2 generate a constant shift from (x,y) to $(\hat{x},\hat{y})$, namely the shift. Noting this allows us to "factor out" the Lie symmetries so that we're down to

$$\hat{x}(x,y) = b_1^1 x + b_2^1 y, \qquad \hat{y}(x,y) = b_1^2 x + b_2^2 y. \tag{63}$$

The remaining symmetries may be generated from the above by using the Lie symmetries. For our problem every discrete symmetry is of the form (63) for some matrix $\boldsymbol{B}$ up to equivalence under translations. Let's substitute (63) into the symmetry condition $\hat{y}'' = \tan\hat{y}'$ when $y'' = \tan y'$.

Using the prolongation formula $\hat{y}^{(1)} = \frac{D_x\hat{y}^{(0)}}{D_x\hat{x}} = \frac{\left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\hat{y}}{\left(\partial_x + y'\partial_y + y''\partial_{y'} + \cdots\right)\hat{x}}$, we get

$$\hat{y}^{(1)} = \hat{y}' = \frac{b_1^2 + b_2^2 y'}{b_1^1 + b_2^1 y'}.$$

With

$$\hat{y}^{(2)} = \frac{D_x\hat{y}^{(1)}}{D_x\hat{x}} = \frac{D_x\left(\frac{b_1^2 + b_2^2 y'}{b_1^1 + b_2^1 y'}\right)}{D_x(b_1^1 x + b_2^1 y)} = \frac{\left(\frac{b_2^1(b_1^1 + b_2^1 y') - b_2^2(b_1^2 + b_2^2 y')}{(b_1^1 + b_2^1 y')^2}\right) y''}{b_1^1 + b_2^1 y'} = \frac{(b_2^1 b_1^1 - b_1^2 b_2^2) y''}{(b_1^1 + b_2^1 y')^3}$$

$$= \frac{\mathrm{Det}(\boldsymbol{B}) y''}{(b_1^1 + b_2^1 y')^3}.$$

Let $J \equiv \mathrm{Det}(\boldsymbol{B})$. Then

$$\hat{y}^{(2)} = \hat{y}'' = \frac{\mathrm{J}y''}{(b_1^1 + b_2^1 y')^3} = \frac{\mathrm{J}\tan y'}{(b_1^1 + b_2^1 y')^3}. \tag{64}$$

Therefore the symmetry condition is:

$$\frac{\mathrm{J}\tan y'}{(b_1^1 + b_2^1 y')^3} = \tan(\hat{y}') = \tan\left(\frac{b_1^2 + b_2^2 y'}{b_1^1 + b_2^1 y'}\right).$$

If we differentiate the above equation wrt y' (noting that $\frac{d\tan y'}{dy'} = \sec^2 y' = 1 + \tan^2 y'$) we get

$$\frac{J(1+\tan^2 y')}{(b_1^1 + b_2^1 y')^3} - \frac{3b_2^1 J\tan y'}{(b_1^1 + b_2^1 y')^4} = \left(\frac{b_2^1(b_1^1 + b_2^1 y') - b_2^2(b_1^2 + b_2^2 y')}{(b_1^1 + b_2^1 y')^2}\right)\left(1 + \tan^2\left(\frac{b_1^2 + b_2^2 y'}{b_1^1 + b_2^1 y'}\right)\right).$$

Cleaning up the RHS we get

$$\frac{J(1+\tan^2 y')}{(b_1^1 + b_2^1 y')^3} - \frac{3b_2^1 J\tan y'}{(b_1^1 + b_2^1 y')^4} = \frac{J}{(b_1^1 + b_2^1 y')^2}\left(1 + \tan^2\left(\frac{b_1^2 + b_2^2 y'}{b_1^1 + b_2^1 y'}\right)\right).$$

Let's clean up this result to

$$(b_1^1 + b_2^1 y')J(1+\tan^2 y') - 3b_2^1 J\tan y' = J(b_1^1 + b_2^1 y')^2\left(1 + \tan^2\left(\frac{b_1^2 + b_2^2 y'}{b_1^1 + b_2^1 y'}\right)\right).$$

Putting the trigonometric terms on the left hand side and the rest on the right hand side we get:

$$(b_1^1 + b_2^1 y')J\left((b_1^1 + b_2^1 y')\tan^2\left(\frac{b_1^2 + b_2^2 y'}{b_1^1 + b_2^1 y'}\right) - \tan^2 y'\right) + 3b_2^1 J\tan y'$$
$$= (b_1^1 + b_2^1 y')J(1 - (b_1^1 + b_2^1 y'))$$

If $b_2^1 \neq 0$. Then we have an algebraic equation for $\tan y'$ in terms of y'. This can't be as the tangent is a transcendental function (it has no finite polynomial representation). This forces us to conclude that $b_2^1 = 0$.

$$Jb_1^1\left(b_1^1\tan^2\left(\frac{b_1^2+b_2^2y'}{b_1^1}\right)-\tan^2 y'\right)=Jb_1^1(1-b_1^1).$$

The problem persists unless $b_1^1=1$. Then

$$\tan^2(b_1^2+b_2^2y')-\tan^2 y'=0.$$

Finally $\tan y'=\tan(b_1^2+b_2^2y')$ if $b_2^2=\alpha\in\{-1,1\}$ and $b_1^2=q\pi x, q\in\mathbb{Z}$. Then (64) reduces to

$$\alpha\tan y'=\tan(q\pi x+\alpha y').$$

Therefore, finally, recalling that we started with $\hat{x}(x,y)=b_1^1x+b_2^1y$, $\hat{y}(x,y)=b_1^2x+b_2^2y$. the inequivalent discrete symmetries are:

$$(\hat{x},\hat{y})=(x,\alpha y+q\pi x),\qquad \alpha\in\{-1,1\},\qquad q\in\mathbb{Z}.$$

If $\mathcal{L}$ is abelian and $R\geq 2$, computer algebra should be used as the number of unknown coefficients b_i^l increases rapidly with R. If $\mathcal{L}$ non-abelian it is possible to factor out Lie symmetries before solving (62). (Often this reduces the number of nonzero coefficients in B from R^2 to R. Then it is possible to find the discrete symmetries of a differential equation even if R is not small simplifying matrix B using essentially the same method used to classify the generators of one-parameter symmetry groups.

Classification of discrete symmetries. Particle and field theorists usually classify symmetries using matrix representations of Lie algebras. If $\mathcal{L}$ is non-abelian then at least some of the commutators $[X_i,X_j]=c_{ij}^kX_k$are nonzero. Therefore the generators belong to nonempty equivalence classes. We may use these equivalence classes to simplify B. Recall $\tilde{X}_i=\left(A(j,\varepsilon)\right)_i^m X_m$ under the Lie symmetries generated by X_j. We can write equation (61), as

$$\tilde{X}_i = \tilde{b}_i^l \hat{X}_l, \tag{65}$$

where $\tilde{b}_i^l = \big(A(j,\varepsilon)\big)_i^p b_p^l$. We see that equation (65) is equivalent to equation (61)

$$X_i = \tilde{b}_i^l \hat{X}_l. \tag{66}$$

Then it must be that the solutions $\hat{z}$ of (65) are related to the solutions of

$$\zeta_i^r(z)\frac{\partial \hat{z}^s}{\partial z^r} = b_i^l \zeta_i^s(\hat{z}), \tag{67}$$

by symmetries in the one-parameter group generated by X_j. That is, $B \sim A(j,\varepsilon)B$, the tilde denoting equivalence in the sense of matrices. To "factor our" the Lie symmetries generated b X_j, solve (67) for just one (simple) matrix in this family.

Continuing with the above motif, the generator $\tilde{X}_l \sim \big(A(j,\varepsilon)\big)_i^m \hat{X}_m$ under the Lie symmetries generated by $\hat{X}_j$. Therefore—using the same argument as above—B is equivalent to $BA(j,\varepsilon)$. An ***equivalence transformation*** describes the replacement of B with $A(j,\varepsilon)B$ or $BA(j,\varepsilon)$.

For abelian Lie algebras the elements of B are unrelated, but this is not so for non-abelian Lie algebras. (Remember that we are finding equivalences between generators that do not commute.) These relationships together with the equivalence transformations usually enable us to reduce R to a greatly simplified form. (Recall that the structure constants (as the name implies) remain unaltered by a change of basis. Then if the generators satisfy $[X_i, X_j] = c_{ij}^k X_k$ they satisfy

$$[\hat{X}_i, \hat{X}_j] = c_{ij}^k \hat{X}_k. \tag{68}$$

Substitute $X_i = b_i^l \hat{X}_l$ into $[X_i, X_j] = c_{ij}^k X_k$to get

$$b_i^l b_j^m [\hat{X}_l, \hat{X}_m] = c_{ij}^k b_k^n \hat{X}_n.$$

Then (68) leads to the identities

$$c_{lm}^{n} b_i^l b_j^m = c_{ij}^{k}.\, b_k^n. \qquad (69)$$

These identities are loaded with physics meaning showing you how quantum operators relate to each other. The equations are nonlinear constraints on B. The constraints with $i \geq j$ are essentially the same as those for $i < j$. The constraints are not affected by an equivalence transformation by the matrices $A(j, \varepsilon)$. The order in which the matrices $A(j, \varepsilon)$ are used does not affect the classification of the matrices B. Any ordering gives the same final form ***provided*** that the parameters ε are chosen appropriately.

<u>Example 2.18</u>—Assume that we have applied the linearized symmetry condition to some ODE or PDE to get a Lie algebra of infinitesimal generators X_1, X_2 such that $[X_1, X_2] = X_1$. The only nonzero structure constants are $c_{12}^1 = 1,\ c_{21}^1 = -1$. The constraints (68) (with $i < j$)) are

$$c_{12}^1 b_1^1 b_2^2 - c_{21}^1 b_1^2 b_2^1 = b_1^1, \qquad (i, j, n) = (1,2,1),$$

$$0 = b_1^2, \qquad (i, j, n) = (1,2,2),$$

and therefore (recalling that B is nonsingular)

$$B = \begin{bmatrix} b_1^1 & 0 \\ b_2^1 & 1 \end{bmatrix}, \qquad b_1^1 \neq 0.$$

As we did in example 2.15, we need to compute the $A(j, \varepsilon) = e^{\varepsilon C(j)}$. We begin with $C(1)$.

$$C(1) = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}.$$

$$A(1, \varepsilon) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\varepsilon}{1!}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{\varepsilon^2}{2!}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}^2 + \cdots = \begin{bmatrix} e^{\varepsilon} & 0 \\ 0 & 1 \end{bmatrix}.$$

$$A(2, \varepsilon) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\varepsilon}{1!}\begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix} + \frac{\varepsilon^2}{2!}\begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}^2 + \cdots = \begin{bmatrix} 1 & 0 \\ -\varepsilon & 1 \end{bmatrix}.$$

Multiplying B on the RHS by $A(1,\varepsilon)$ leading to: $BA(1,\varepsilon)=\begin{bmatrix} b_1^1 & 0 \\ b_2^1-\varepsilon & 1\end{bmatrix}$. Let $b_2^1=\varepsilon$ reducing $BA(1,\varepsilon)$ to $BA(1,\varepsilon)=\begin{bmatrix} b_1^1 & 0 \\ 0 & 1\end{bmatrix}$. Now do $BA(2,\varepsilon)$ to get $BA(1,\varepsilon)=\begin{bmatrix} b_1^1e^{\varepsilon} & 0 \\ 0 & 1\end{bmatrix}$. Let $e^{\varepsilon}=-\ln\left|b_1^1\right|$. This is the same as letting $b_1^1=1$ or $b_1^1=-1$. Let's check this. Let $b_1^1=1$. Then $b_1^1e^{\varepsilon}=b_1^1e^{-\ln\left|b_1^1\right|}=b_1^1e^{\ln\left|\frac{1}{b_1^1}\right|}=\frac{b_1^1}{b_1^1}=1$. Let $b_1^1=-1$. Then $-b_1^1>0$. So $b_1^1e^{\varepsilon}=b_1^1e^{-\ln\left|-b_1^1\right|}=b_1^1e^{-\ln\left|b_1^1\right|}=b_1^1e^{\ln\left|\frac{1}{b_1^1}\right|}=\frac{b_1^1}{b_1^1}=1$. We cannot simplify B any more. Thus$B=\begin{bmatrix} \alpha & 0 \\ 0 & 1\end{bmatrix}, \alpha=\{-1,1\}$.

<u>Example 2.19</u>—A physics rich example (see $\mathrm{SL}_2(\mathbb{R})$ http://en.wikipedia.org/wiki/SL2%28R%29 . Consider $\mathfrak{sl}(2)$. There are many ODEs and PDEs whose Lie algebra (extracted from the linearized symmetry condition) is $\mathfrak{sl}(2)$ It's structure constants are $c_{12}^1=-c_{21}^1=1$, $c_{13}^2=-c_{31}^2=2$, $c_{23}^3=-c_{32}^3=1$. Go back to example 2.15 for all of the details.

$$A(1,\varepsilon)=\begin{bmatrix} 1 & 0 & 0 \\ -\varepsilon & 1 & 0 \\ \varepsilon^2 & -2\varepsilon & 1\end{bmatrix},\qquad A(2,\varepsilon)=\begin{bmatrix} e^{\varepsilon} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-\varepsilon}\end{bmatrix},\qquad A(3,\varepsilon)=\begin{bmatrix} 1 & 2\varepsilon & \varepsilon^2 \\ 0 & 1 & \varepsilon \\ 0 & 0 & 1\end{bmatrix}.$$

The constraints (68), $c_{lm}^n b_i^l b_j^m=c_{ij}^k.b_k^n$, are highly coupled because $\mathfrak{sl}(2)$ is a simple Lie algebra (its only ideals are itself and the set $\{0\}$). For example take $n=1$. Then

$$b_1^1b_2^2-b_1^2b_2^1=b_1^1, \tag{70}$$

$$b_1^1b_3^2-b_1^2b_3^1=2b_2^1, \tag{71}$$

$$b_2^1b_3^2-b_2^2b_3^1=b_3^1. \tag{72}$$

(It's maddening to read this stuff in a physics text or paper and have no real context like we have.) If $b_1^1\neq 0$, multiply B on the LHS by $A(1,b_2^1/b_1^1)$, which is equivalent to setting $b_2^1=0$. Then (70) gives $b_2^2=1$, and so (72) is satisfied. As a consequence (71) yields $b_3^2=0$. Our simplified matrix B looks like

$$B=\begin{bmatrix} b_1^1 & b_1^2 & b_1^3 \\ 0 & 1 & b_2^3 \\ 0 & 0 & b_3^3 \end{bmatrix}.$$

Multiply this Bon the RHS by $A(3,-b_1^2/2b_1^1)$ to set the (1,2) matrix element to zero. Then the remaining constraints (68), $c_{lm}^n b_i^l b_j^m = c_{ij}^k . b_k^n$ can be satisfied only if $b_1^3 = b_2^3 = 0$ and $b_3^3 = 1/b_1^1$. Lastly, multiply B by $A(2,-\ln|b_1^1|$ to get two inequivalent matrices:

$$B=\begin{bmatrix} \alpha & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \alpha \end{bmatrix}, \qquad \alpha=\{-1,1\}.$$

The last possibility is that $b_1^1 = 0$. Working out the details leads to

$$B=\begin{bmatrix} 0 & 0 & \alpha \\ 0 & -1 & 0 \\ \alpha & 0 & 0 \end{bmatrix}, \qquad \alpha=\{-1,1\}.$$

For $\mathfrak{sl}(2)$ there are four distinct matrices B that are inequivalent under Lie symmetries.

<u>Example 2.20</u>—(See Hydon text). The Chazy equation is $y''' = 2yy'' - 3y'^2 + \lambda(6y' - y^2)^2$. If you work out the details there are only two real-valued discrete point symmetries (up to equivalence):

$$(\hat{x},\hat{y}) = \{(x,y),(-x,y)\}.$$

Here is what I got on the Chazy equation from the internet: a deep connection to mathematical physics. The general solution to the Chazy equation can be expressed as the ratio of two solutions to a hypergeometric equation. The reduction method leads to an alternative formula in terms of solutions to the Lamé equation, resulting in a surprising transformation between the Lamé and hypergeometric equations.

<u>Example 2.21</u>— (See Hydon text) In mathematics, and in particular in the theory of solitons, the **Dym equation** (**HD**) is the third-order partial differential equation

$$u_t = u^3 u_{xxx}.$$

From the application of the linearized symmetry to the HD, we discover that it has a five-dimensional Lie algebra. The basis are

$$X_1 = \partial_x, \qquad X_2 = x\partial_x + u\partial_u, \qquad X_3 = x^2\partial_x + 2xu\partial_u, \qquad X_4 = \partial_t, \qquad X_5 = t\partial_t - \frac{1}{3}u\partial_u.$$

The Dym equation represents a system in which dispersion and nonlinearity are coupled together. HD is a completely integrable nonlinear evolution equation that may be solved by means of the inverse scattering transform. It is interesting because it obeys an infinite number of conservation laws (Wow!); it does not possess the Painlevé property. A friend of mine thinks particle and field theorists pay too much attention to theories that can be linearized and treated by (Feynman) perturbation methods. I have a feeling it would take me a long time, months perhaps, to get any kind of real understanding of all that is going on with HD equation. There are eight inequivalent real discrete symmetries:

$$(\hat{x}, \hat{t}, \hat{u}) = \left\{(\alpha x, \beta t, \alpha\beta u), \left(-\frac{\alpha}{x}, \beta t, \frac{\alpha\beta u}{x^2}\right)\right\}, \qquad \alpha, \beta \in \{-1,1\}.$$

I am exhausted. When I was a PhD student at the University of Houston Department of Physics, there was a Professor Golubitsky who offered classes in areas I am only now just getting enough foundations in. He is an expert in bifurcation theory, which describes the way in which nonlinear systems change their behavior as parameters are varied. For systems with symmetries, equivalent bifurcation theory is needed to deal with degeneracies that are associate with the symmetries (see Golubitsky, Stewart, and Schaeffer (1988)).

<u>Part I</u>. **Chapter 3**. (Accessible to sophomores; required for mathematics/physics majors up through the postdoctoral research level) On to higher order ODEs.

The last practical example from chapter two was example 2.8. If you're the sophomore who's just finished a first course in ODEs, it's fine to skip from that example to this chapter without any loss of continuity for we continue with solving ODEs. You cannot, however, be a master of differential equations or a true physicist without learning the classification of continuous and discrete symmetries covered in the latter parts of chapter two, and I'm not just talking about stuff important to foundation physics like particles and fields. I'm talking about having a clue as to how systems change behavior across discrete symmetries (Golubitsky, Stewart, and Schaeffer (1988)). If you've plowed through chapter two, a quick review of the first part of chapter two up through example 2.8 is good advice. We are going back to example 2.3 and reducing the second order ODE to an algebra equation; we had reduced it into two different first order ODEs.

Chapters 1 and 2 have deeply supplemented and rendered Hydon chapters 1-5, 10 and 11 highly transparent and self-contained. Here are the remaining Hydon text goals.

1-Review the remaining material on ODE symmetry methods in chapters 6 and 7 of Hydon's text. The step-by-step notes from chapter 1 through example 2.8 have given you every single tool necessary to deal with this material. Thus I will only outline this material.

2-Show how symmetry methods for ODEs extend to PDEs (Hydon chapters 8 and 9).

3-Wrap up a few miscellaneous Hydon sections, most importantly dealing with variational symmetries (important to physicists).

Example 3.1 from 2.3—Consider the nonlinear ODE $y'' = \frac{y'^2}{y} - y^2$. Back in example 2.3 we applied the linearized symmetry condition to get the Lie point symmetries

$$X_1 = \partial_x, \qquad X_2 = x\partial_x - 2y\partial_y.$$

In reducing the ODE to a first order ODE using X_1 we had derived the differential invariants

$$r_1 = y, \;\; v_1 = y'$$

We used X_2 to derive $r_2 = y'y^{-\frac{3}{2}}$, but I fed you v_2. If you apply the steps of example 2.7 to this example, you get

$$v_2 = \frac{y''}{y^2}.$$

The algebra in example 2.3 then shows $v_2 = \frac{y''}{y^2} = \frac{y'^2}{y^3} - 1 = \frac{y' \cdot y'}{y^{\frac{3}{2}} \cdot y^{\frac{3}{2}}} - 1 = r_2^2 - 1$. That is, $v_2 = r_2^2 - 1$.

As you can verify from example 2.3, the generators X_1, X_2 form a solvable Lie algebra. These generators are, moreover, a canonical basis (see 2nd half of chapter 2). Thus X_2 generates point transformations of the variables (r_1, v_1). Explicitly,

$$X_2 r_1 = -2y = -2r_1, \qquad X_2^{(1)} v_1 = -3y' = -3v_1.$$

(I derived the prolonged generator $X_2^{(1)}$ in example 2.3. It is $X_2^{(1)} = x\partial_x - 2y\partial_y - 3y'\partial_{y'}$.) Thus the ***restriction*** (new stuff) of X_2 to (r_1, v_1) is $X_2 = -2r_1\partial_{r_1} - 3v_1\partial_{v_1}$.

We've already chosen the invariant canonical coordinate $r_2 = \frac{v_1}{r_1^{3/2}}$. Let $s_2 = -\frac{1}{2}\ln(r_1)$. Then

$$\frac{ds_2}{dr_2} = \frac{\frac{ds_2}{dx}}{\frac{dr_2}{dx}} = \frac{r_2}{3r_2^2 - 2v_2}.$$

From $v_2 = r_2^2 - 1$ we see that

$$\frac{ds_2}{dr_2} = \frac{r_2}{r_2^2 + 2}.$$

Quadrature results in $s_2 = \frac{1}{2}\ln(r_2^2 + 2) + c$. Rewrite this solution in terms of (r_1, v_1) to get the algebraic equation

$$v_1 = \pm r_1(4c_1^2 - 2r_1)^{\frac{1}{2}} + c.$$

We have completed one iteration. Now we do another iteration, now using the generator X_1 to solve our algebraic equation. With the canonical coordinates $(r, s) = (y, x)$, we obtain

$$\frac{ds_1}{dr_1} = \frac{1}{y'} = \pm \frac{1}{r_1(4c_1^2 - 2r_1)^{\frac{1}{2}}}.$$

Therefore the general solution to our ODE is

$$s_1 = c_2 \mp c_1^{-1} \cosh\left(c_1 \sqrt{\frac{2}{r_1}} \right).$$

Back in our original variables we get

$$y = 2c_1^2 \operatorname{sech}^2\big(c_1(x - c_2)\big).$$

The biggest take away is the restriction of generators to particular canonical coordinates.

Please refer to Hydon's text for the next concept: <u>The order which you use the generators can short circuit the iteration process</u>. In example 6.2 and 6.3 Hydon treats the ODE

$$y''' = \frac{y''^2}{y'(1 + y')}$$

Iteratively, first with Lie generators ordered as $X_1 = \partial_x,\ X_2 = \partial_y,\ X_3 = x\partial_x + y\partial_y$, then ordered as $X_1 = \partial_y,\ X_2 = \partial_x,\ X_3 = x\partial_x + y\partial_y$. The first ordering leads to a solution, the second ordering leads to a dead end. Order matters.

A useful high school algebra tip. Sometimes you might get a system of algebraic equations for your differential invariants such as

$$v^3 - 2(r - c_1)v + 4 = 0$$

$$v^3 - 2(s - c_2)v^2 + 1 = 0.$$

As you would with Gauss-Jordon elimination, subtract the bottom equation from the top equation to get

$$v^3 - 2(r - c_1)v + 4 = 0$$

$$2(s - c_2)v^2 - 2(r - c_1)v + 3 = 0.$$

Multiply the lower equation by an appropriate power of v to eliminate v^3 from the top equation to get

$$2(r - c_1)v^2 - \{2(r - c_1)(s - c_2) + 3\}v + 4(s - c_2) = 0$$

$$2(s - c_2)v^2 - 2(r - c_1)v + 3 = 0.$$

Iterate until a solution for v is found:

$$v = \frac{6(r - c_1) - 4(s - c_2)^2}{4(r - c_1)^2 - 2(r - c_1)(s - c_2)^2 - 3(s - c_2)}$$

provided the denominator is nonzero.

New symmetries obtained during reduction (Important). In some cases the Lie algebra of a differential equation is sufficiently large to enable us to solve the differential equation completely. This needn't be the case. Let us consider solving ODEs of order n whose largest solvable subalgebras are of order $n-1$ or less. Let $\{X_1, \ldots, X_S\}$ be a canonical basis for such a subalgebra. As we've learned in chapter 2 in terms of the fundamental differential invariants (r_S, s_S) of this subalgebra, the ODE is equivalent to an ODE of order $n-S$. The general solution of the reduced ODE is an algebraic equation

$$v_S = F(r_S; c_1, \ldots, c_{n-S}). \quad (1)$$

If we can get to this solution we are done because (1) is equivalent to an ODE of order S that admits the symmetries generated by $\{X_1, \ldots, X_S\}$. If we can't solve (1) we might have options. Each subalgebra $\mathcal{L}_k = \text{Span}(X_1, \ldots, X_k)$ in the solvable chain can be used to reduce the order of the ODE to an equivalent ODE of order $n-k$ in the fundamental differential invariants (r_k, v_k). So there is a sequence of intermediate reduced ODEs, and it may be that one of the intermediate ODEs has **NEW** point symmetries, as well as those inherited from the original ODE. With a sufficient number of new symmetries, we might obtain the general solution of an intermediate equation in the form $v_k = F(r_S; c_1, \ldots, c_{n-k})$. Then the symmetries in $\mathcal{L}_k$ may be used to complete the solution of the original ODE.

<u>Example 3.2</u>—Consider $y''' = \frac{2y''^2}{y'} + \frac{y''}{x} + \frac{y'^2}{x}$. The linearized symmetry condition leads us to the two-dimensional Lie algebra generated by $X_1 = \partial_y,\ X_2 = x\partial_x$. The fundamental differential invariants are

$$r_1 = x,\ \ v_1 = y' \text{ for } \mathcal{L}_1 = \text{Span}(X_1)$$

$$r_2 = xy',\ \ v_2 = x^2 y'' \text{ for } \mathcal{L}_2 = \text{Span}(X_2)$$

Choosing the second pair of canonical coordinates and cranking the tools we've learned (and no new stuff), we would find out that our ODE is equivalent to $\frac{dv_2}{dr_2} = \frac{2v_2^2 + 3r_2v_2 + r_2^2}{r_2(v_2 + r_2)}$. It's symmetries are not obvious. However, in terms of the first pair of canonical coordinates, our ODE is also equivalent to the 2nd order ODE $\frac{d^2v_1}{dr_1^2} = \frac{2}{v_1}\left(\frac{dv_1}{dr_1}\right)^2 + \frac{1}{r^2}\left(\frac{dv_1}{dr_1} + v_1^2\right)$, which has an eight-dimensional Lie algebra (**this is the BIG point**). The symmetries generated from X_2 are inherited form the original ODE, but the remaining (larger set) of symmetries are new. **These new symmetries** (see Hydon example 6.5) **lead us to the solution of our ODE**. (These notes give you all the step-by-step tools to work out this problem for yourselves.)

Section 6.3 of Hydon is dedicated to a potentially deep example—the integration of third-order ODEs with the Lie algebra $\mathfrak{sl}(2)$. Hydon present nothing new in this example, but it is interesting because shows that the reduced ODE is our famous Riccati equation which can be transformed into the Schrödinger equation. Is there something deep to this connection between Riccati and Schrödinger?

Hydon's chapter seven is lengthy, connecting two of our ideas into a new toolset for ODEs, the Lie generators we compute for an ODE via the linearized symmetry condition to the method of characteristics to find first integrals. It's useful for ODEs whose Lie algebra is not solvable. The method covered in Hydon section 7.1 is limited by the need for at least $n+1$ generators. Section 7.2 introduces contact symmetries and dynamical symmetries. They are defined as follows. A diffeomorphism

$$(\hat{x},\hat{y},\hat{y}')=\left(\hat{x}(x,y,y'),\hat{y}\big(\hat{x}(x,y,y')\big),\hat{y}'(x,y,y')\right)$$

is a contact transformation if

$$\hat{y}'(x,y,y')=\frac{\hat{y}_x+y'\hat{y}_y+y''\hat{y}_{y'}}{\hat{x}_x+y'\hat{x}_y+y''\hat{x}_{y'}}.$$

You can supplement material on contact transformations by looking them up in Herbert Goldstein's graduate text on classical mechanics. Any generator X whose characteristic Q satisfies the linearized symmetry condition generates ***dynamical symmetries*** (or **internal symmetries** (does this mean the same as internal symmetries in quantum field theories?)). Finally Hydon section 7.3 relates integrating factors (studied in regular botany-based ODEs) to the material in section 7.1 Why the rush?

How to obtain Lie point symmetries of PDEs. (A junior level course in PDEs is helpful.) There are no new big ideas when extending from symmetry methods for ODEs to PDEs. We only deal with more indices. Let's begin with scalar PDEs with one dependent variable u and two independent variables x and t. We express an n^th order ODE by

$$\Delta(x, t, u, u_x, u_t, \dots) = 0.$$

We restrict ourselves for now for PDEs of the form

$$\Delta = u_\sigma - \omega(x, t, u, u_x, u_t, \dots) = 0,$$

where u_σ is one of the n^th order derivatives of u and ω is independent of u_σ. It could be that u_σ is of order $n - k$ provided ω is independent of u_σ or any derivatives of u_σ. A point transformation is a diffeomorphism

$$\Gamma: (x, t, u) \longrightarrow \Big(\hat{x}(x, t, u), \hat{t}\big(\hat{x}(x, t, u)\big), \hat{u}(x, t, u)\Big).$$

It maps the surface $u = u(x, t)$ to the surface

$$\hat{x} = \hat{x}\big(x, t, u(x, t)\big),$$

$$\hat{t} = \hat{t}\big(x, t, u(x, t)\big),$$

$$\hat{u} = \hat{u}\big(x, t, u(x, t)\big),$$

Our goal, namely the symmetry condition for PDEs, is that we want

$$\Delta(\hat{x}, \hat{t}, \hat{u}, \hat{u}_{\hat{x}}, \hat{u}_{\hat{t}}, \dots) = 0$$

when

$$\Delta(x, t, u, u_x, u_t, \dots) = 0.$$

If this is so, then Γ is a point symmetry of $\Delta(x, t, u, u_x, u_t, \dots) = 0$. To see when this goal is attainable we have to know under what conditions we may invert the equations for $\hat{x}$ and $\hat{t}$ (at least locally) to give x and t in terms of $\hat{x}$ and $\hat{t}$, and similarly for all of the partial derivatives of u and their counterparts $\hat{u}$.

What binds the two coordinate systems together is the total derivative, one for each independent variable. In our restricted case the total derivatives are

$$D_x = \partial_x + u_x\partial_u + u_{xx}\partial_{u_x} + u_{xt}\partial_{u_t} + \cdots,$$

$$D_t = \partial_t + u_t\partial_u + u_{tt}\partial_{u_t} + u_{xt}\partial_{u_t} + \cdots.$$

As you can see, total derivatives treat the dependent variable u and its derivatives as functions of the independent variables.

If the Jacobian

$$\mathfrak{J} = \begin{vmatrix} D_x\hat{x} & D_x\hat{t} \\ D_t\hat{x} & D_t\hat{t} \end{vmatrix} \neq 0 \tag{2}$$

when $u = u(x, t)$, then we may invert the equations for $\hat{x}$ and $\hat{t}$ (at least locally) to give x and t in terms of $\hat{x}$ and $\hat{t}$. If the Jacobian $\mathfrak{J} \neq 0$, we may rewrite

$$\hat{u} = \hat{u}(\hat{x}, \hat{t}). \tag{3}$$

Applying the chain rule to (3) leads us to

$$\begin{bmatrix} D_x\hat{u} \\ D_t\hat{u} \end{bmatrix} = \begin{bmatrix} D_x\hat{x} & D_x\hat{t} \\ D_t\hat{x} & D_t\hat{t} \end{bmatrix} \begin{bmatrix} \hat{u}_x \\ \hat{u}_t \end{bmatrix}.$$

By Cramer's rule, then

$$\hat{u}_{\hat{x}} = \frac{1}{\mathfrak{J}} \begin{vmatrix} D_x\hat{u} & D_x\hat{t} \\ D_t\hat{u} & D_t\hat{t} \end{vmatrix}, \qquad \hat{u}_t = \frac{1}{\mathfrak{J}} \begin{vmatrix} D_x\hat{x} & D_x\hat{u} \\ D_t\hat{x} & D_t\hat{u} \end{vmatrix}.$$

Higher order prolongations to higher order partial derivatives are obtained recursively by repeating the above argument. If we let $\hat{u}_J$ be any derivative of $\hat{u}$ wrt $\hat{x}$ and $\hat{t}$, then

$$\hat{u}_{J\hat{x}} \equiv \frac{\partial \hat{u}_J}{\partial \hat{x}} = \frac{1}{\mathfrak{J}} \begin{vmatrix} D_x\hat{u}_J & D_x\hat{t} \\ D_t\hat{u}_J & D_t\hat{t} \end{vmatrix},$$

$$\hat{u}_{J\hat{t}} \equiv \frac{\partial \hat{u}_J}{\partial \hat{t}} = \frac{1}{\mathfrak{J}} \begin{vmatrix} D_x\hat{x} & D_x\hat{u}_J \\ D_t\hat{x} & D_t\hat{u}_J \end{vmatrix}.$$

For example, the transformation is prolonged to second derivatives as follows:

$$\hat{u}_{\hat{x}\hat{x}} = \frac{1}{\mathfrak{J}} \begin{vmatrix} D_x\hat{u}_{\hat{x}} & D_x\hat{t} \\ D_t\hat{u}_{\hat{x}} & D_t\hat{t} \end{vmatrix}, \qquad \hat{u}_{\hat{t}\hat{t}} = \frac{1}{\mathfrak{J}} \begin{vmatrix} D_x\hat{x} & D_x\hat{u}_{\hat{t}} \\ D_t\hat{x} & D_t\hat{u}_{\hat{t}} \end{vmatrix}, \qquad \hat{u}_{\hat{x}\hat{t}} = \frac{1}{\mathfrak{J}} \begin{vmatrix} D_x\hat{u}_{\hat{t}} & D_x\hat{u}_{\hat{t}} \\ D_t\hat{u}_{\hat{t}} & D_t\hat{u}_{\hat{t}} \end{vmatrix} = \frac{1}{\mathfrak{J}} \begin{vmatrix} D_x\hat{x} & D_x\hat{u}_{\hat{x}} \\ D_t\hat{x} & D_t\hat{u}_{\hat{x}} \end{vmatrix}$$

Given all of this, the point transformation Γ is a point symmetry of our PDE if

$$\Delta(\hat{x}, \hat{t}, \hat{u}, \hat{u}_{\hat{x}}, \hat{u}_{\hat{t}}, \dots) = 0$$

when

$$\Delta(x, t, u, u_x, u_t, \dots) = 0.$$

This symmetry is typically extremely complicated so we will linearize it just as we did for ODEs. Before we do this linearization, it is worth working an easy example.

<u>**Example 3.3**</u>—I claim that

$$(\hat{x}, \hat{t}, \hat{u}) = \left(\frac{x}{2t}, \frac{-1}{4t}, 2(ut - x)\right)$$

is a point symmetry of Burger's equation $u_t + uu_x = u_{xx}$. The Jacobian of the point transformation is

$$\mathfrak{J} = \begin{vmatrix} \frac{1}{2t} & 0 \\ \frac{-x}{2t^2} & \frac{1}{4t^2} \end{vmatrix} = \frac{1}{8t^3},$$

Thus

$$\hat{u}_{\hat{x}} = 8t^3 \begin{vmatrix} 2(u_x t - 1) & 0 \\ 2(tu_t + u) & \frac{1}{4t^2} \end{vmatrix} = 4t(u_x t - 1),$$

$$\hat{u}_t = 8t^3 \begin{vmatrix} \frac{1}{2t} & 2(tu_x - 1) \\ \frac{-x}{2t^2} & 2(tu_t + u) \end{vmatrix} = 8t(t^2 u_t + xtu_x + tu - x),$$

$$\hat{u}_{\hat{x}\hat{x}} = 8t^3 \begin{vmatrix} 4t^2 u_{xx} & 0 \\ 4(t^2 u_{xt} + 2tu_x - 1) & \frac{1}{4t^2} \end{vmatrix} = 8t^3 u_{xx}.$$

Then $\hat{u}_{\hat{x}\hat{x}} = \hat{u}_t + \hat{u}\hat{u}_{\hat{x}} = 8t^3(u_t + uu_x) = 8t^3 u_{xx}$. Indeed then our point transformation satisfies the symmetry condition of $\hat{u}_{\hat{x}\hat{x}} = \hat{u}_t + \hat{u}\hat{u}_{\hat{x}}$ when $u_{xx} = u_t + uu_x$.

We cannot in general pull point symmetries to PDEs out of our arses. As with ODEs, we need tools to seek out one-parameter Lie groups of point symmetries. By seeking symmetries of the form

$$\begin{aligned} \hat{x} &= x + \varepsilon\xi(x,t,u) + O(\varepsilon^2), \\ \hat{t} &= t + \varepsilon\xi(x,t,u) + O(\varepsilon^2), \qquad (4) \\ \hat{u} &= u + \varepsilon\xi(x,t,u) + O(\varepsilon^2), \end{aligned}$$

we impose a set of constraints which we can exploit to tease out point symmetries. Just as for Lie point transformations for ODEs (transformations on the plane), each one-parameter, local Lie group of point transformations is obtained by exponentiating its infinitesimal generator

$$X = \xi\partial_x + \tau\partial_t + \eta\partial_u.$$

Equivalently, we can obtain $(\hat{x}, \hat{t}, \hat{u})$ by solving

$$\frac{d\hat{x}}{d\varepsilon} = \xi(\hat{x}, \hat{t}, \hat{u}), \qquad \frac{d\hat{t}}{d\varepsilon} = \tau(\hat{x}, \hat{t}, \hat{u}), \qquad \frac{d\hat{u}}{d\varepsilon} = \eta(\hat{x}, \hat{t}, \hat{u})$$

subject to the initial conditions $(\hat{x}, \hat{t}, \hat{u})_{\varepsilon=0} = (x, t, u)$.

A surface $u = u(x,t)$is mapped to itself by the group of transformations generated by Xif

$$X\big(u - u(x,t)\big) = 0 \tag{5}$$

when $u = u(x,t)$. This condition may be expressed using the characteristic of the group,

$$Q = \eta - \xi u_x - \tau u_t.$$

From (5), the surface $u = u(x,t)$ is invariant provided that $Q = 0$ when $u = u(x,t)$. The equation for Q is called the invariant surface condition, a central equation to some of the main techniques for finding exact solutions to PDEs. Note: Equation (5) seems stupid. However, looking back at the introduction of Q for ODEs, the equation for Q makes sense for PDEs by analogy to ODEs.

The prolongation of the point transformation (4) to first derivatives is

$$\hat{u}_{\hat{x}} = u_x + \eta^x(x,t,u,u_x,u_t) + O(\varepsilon^2), \tag{6}$$
$$\hat{u}_{\hat{t}} = u_t + \eta^t(x,t,u,u_x,u_t) + O(\varepsilon^2).$$

We need explicit formulae for η^x and η^t. Let's derive the formula for η^x in full detail. To get η^x plug in the three equations of (4) up to first order into $\Im = \begin{vmatrix} D_x\hat{x} & D_x\hat{t} \\ D_t\hat{x} & D_t\hat{t} \end{vmatrix}$, and into $\hat{u}_{\hat{x}} = \frac{1}{\Im}\begin{vmatrix} D_x\hat{u} & D_x\hat{t} \\ D_t\hat{u} & D_t\hat{t} \end{vmatrix}$. Recall:

$$D_x = \partial_x + u_x\partial_u + u_{xx}\partial_{u_x} + u_{xt}\partial_{u_t} + \cdots,$$

$$D_t = \partial_t + u_t\partial_u + u_{tt}\partial_{u_t} + u_{xt}\partial_{u_t} + \cdots.$$

$$\Im = \begin{vmatrix} D_x\hat{x} & D_x\hat{t} \\ D_t\hat{x} & D_t\hat{t} \end{vmatrix} = \begin{vmatrix} D_x\big(x + \varepsilon\xi(x,t,u)\big) & D_x\big(t + \varepsilon\xi(x,t,u)\big) \\ D_t\big(x + \varepsilon\xi(x,t,u)\big) & D_t\big(t + \varepsilon\xi(x,t,u)\big) \end{vmatrix} = \begin{vmatrix} 1 + \varepsilon\xi_x + \varepsilon u_x\xi_u & \varepsilon\xi_t + \varepsilon u_t\tau_u \\ \varepsilon\xi_t + \varepsilon u_t\xi_u & 1 + \varepsilon\tau_t + \varepsilon u_t\tau_u \end{vmatrix}$$
$$= (1 + \varepsilon\xi_x + \varepsilon u_x\xi_u)(1 + \varepsilon\tau_t + \varepsilon u_t) - (\varepsilon\xi_t + \varepsilon u_t\xi_u)(\varepsilon\xi_t + \varepsilon u_t\tau_u).$$

Dropping all terms with of order ε^2 or higher we get $\Im = 1 + \varepsilon(\tau_t + u_t\tau_u + \xi_x + u_x\xi_u)$. Now

$$\hat{u}_{\hat{x}} = \frac{1}{\mathfrak{J}}\begin{vmatrix} D_x\big(u+\varepsilon\xi(x,t,u)\big) & D_x\big(t+\varepsilon\xi(x,t,u)\big) \\ D_t\big(u+\varepsilon\xi(x,t,u)\big) & D_t\big(t+\varepsilon\xi(x,t,u)\big)\end{vmatrix} = \frac{1}{\mathfrak{J}}\begin{vmatrix} u_x+\varepsilon\eta_x+\varepsilon u_x\eta_u & \varepsilon\tau_x+\varepsilon u_x\tau_u \\ u_t+\varepsilon\eta_t+\varepsilon u_x\eta_y & 1+\varepsilon\tau_t+\varepsilon u_t\tau_u\end{vmatrix}$$

$$= \frac{1}{\mathfrak{J}}(u_x+\varepsilon\eta_x+\varepsilon u_x\eta_u+\varepsilon u_x\tau_t+\varepsilon u_x u_t\tau_u-\varepsilon u_t\tau_x-\varepsilon u_t u_x\tau_u).$$

So,

$$\hat{u}_{\hat{x}} = \frac{u_x+\varepsilon\eta_x+\varepsilon u_x\eta_u+\varepsilon u_x\tau_t-\varepsilon u_t\tau_x}{1+\varepsilon(\tau_t+u_t\tau_u+\xi_x+u_x\xi_u)}.$$

Recall $\frac{1}{1+x}\approx 1-x$ for small x. Thus

$$\begin{aligned}\hat{u}_{\hat{x}} = \big(u_x+\varepsilon(\eta_x+u_x\eta_u+u_x\tau_t-u_t\tau_x)\big)\big(1-\varepsilon(\tau_t+u_t\tau_u+\xi_x+u_x\xi_u)\big)\\ = u_x+\varepsilon\eta_x+\varepsilon u_x\eta_u+\varepsilon u_x\tau_t-\varepsilon u_t\tau_x-\varepsilon u_x\tau_t-\varepsilon u_x u_t\tau_u-\varepsilon u_x\xi_x-u_x^2\xi_u\\ = u_x+\varepsilon\eta_x+\varepsilon u_x\eta_u-\varepsilon u_t\tau_x-\varepsilon u_x u_t\tau_u-\varepsilon u_x\xi_x-u_x^2\xi_u\\ = u_x+\varepsilon(\eta_x+u_x\eta_u)-\varepsilon u_x(\xi_x+u_x\xi_u)-\varepsilon u_t(\tau_x+u_x\tau_u)\\ = u_x+\varepsilon(D_x\eta-u_xD_x\xi-u_tD_t\tau).\end{aligned}$$

Since to first order $\hat{u} = u+\varepsilon\xi(x,t,u)$, we see that

$$\eta^x(x,t,u,u_x,u_t) = D_x\eta-u_xD_x\xi-u_tD_x\tau.$$

If you repeat all of this work using the Jacobian and $\hat{u}_t = \frac{1}{\mathfrak{J}}\begin{vmatrix} D_x\hat{x} & D_x\hat{u} \\ D_t\hat{x} & D_t\hat{u}\end{vmatrix}$, you will arrive at:

$$\eta^t(x,t,u,u_x,u_t) = D_t\eta-u_xD_t\xi-u_tD_t\tau.$$

We use recursion to prolong the transformation to higher order derivatives. Suppose that

$$\hat{u}_J = u_J+\varepsilon\eta^J+O(\varepsilon^2), \qquad u_J = \frac{\partial^{j_1+j_2}u}{\partial x^{j_1}\partial t^{j_2}}, \qquad \hat{u}_J = \frac{\partial^{j_1+j_2}\hat{u}}{\partial\hat{x}^{j_1}\partial\hat{t}^{j_2}}$$

for some integers j_1, j_2. Then (just as for η^x and η^t) with

$$\hat{u}_{J\hat{x}} \equiv \frac{\partial \hat{u}_J}{\partial \hat{x}} = \frac{1}{\mathfrak{J}}\begin{vmatrix} D_x\hat{u}_J & D_x\hat{t} \\ D_t\hat{u}_J & D_t\hat{t}\end{vmatrix}, \qquad \hat{u}_{J\hat{t}} \equiv \frac{\partial \hat{u}_J}{\partial \hat{t}} = \frac{1}{\mathfrak{J}}\begin{vmatrix} D_x\hat{x} & D_x\hat{u}_J \\ D_t\hat{x} & D_t\hat{u}_J\end{vmatrix}.$$

We get

$$u_{J\hat{x}} = D_{Jx} + \varepsilon\eta^{Jx} + O(\varepsilon^2),$$

$$u_{J\hat{t}} = D_{Jt} + \varepsilon\eta^{Jt} + O(\varepsilon^2),$$

where

$$\eta^{Jx} = D_x\eta^J - u_{Jx}D_x\xi - u_{Jt}D_x\tau,$$

$$\eta^{Jt} = D_t\eta^J - u_{Jx}D_t\xi - u_{Jt}D_t\tau.$$

Alternatively, we can express the functions η^J in terms of the characteristic, *e.g.*,

$$\eta^x = D_xQ + \xi u_{xx} + \tau u_{xt,}$$

$$\eta^t = D_tQ + \xi U_{xt} + \tau U_{tt}.$$

The higher order terms are obtained by induction on j_1 and j_2:

$$\eta^J = D_JQ + \xi D_J u_x + \tau D_J u_{t,}$$

where $D_J = D_x^{j_1}D_t^{j_2}$.

The infinitesimal generator is prolonged to derivatives by adding all terms of the form $n^J\partial_{u_J}$ up to the desired order. For example,

$$X^{(1)} = \varepsilon\partial_x + \tau\partial_t + \eta\partial_u + \eta^x\partial_{u_x} + \eta^t\partial_{u_t} = X + \eta^x\partial_{u_x} + \eta^t\partial_{u_t},$$

$$X^{(2)} = X^{(1)} + \eta^{xx}\partial_{u_{xx}} + \eta^{xt}\partial_{u_{xt}} + \eta^{tt}\partial_{u_{tt}}.$$

For simplicity, we adopt the notation that the generator is understood to be prolonged as much as necessary to describe the group's action on all of the variables. To find Lie point symmetries, we need explicit expressions for η^{Jx} and η^{Jt}. Here are some (you know how to derive them). The first one is a cut and paste and slight rearrangement from the detailed derivation of $\hat{u}_{\hat{x}}$:

$$\eta^x = \eta_x + (\eta_u - \xi_x)u_x - \tau_x u_t - \xi_u u_X^2 - \tau_u u_x u_T$$

$$\eta^t = \eta_t - \xi_t u_x + (\eta_u - \xi_x)u_t - \xi_u u_x u_t - \tau_u u_t^2,$$

$$\eta^{xx} = \eta_{xx} + (2\eta_{xu} - \xi_{xx})u_x - \tau_{xx}u_t + (\eta_{uu} - 2\xi_{xu})u_X^2 - 2\tau_{xu}u_x u_t - \xi_{uu}u_X^3 - \tau_{uu}u_X^2 u_t$$
$$+ (\eta_u - 2\xi_x)u_{xx} - 2\tau_x u_{xt} - 3\xi_u u_x u_{xx} - \tau_u u_t u_{xx} - 2\tau_u u_x u_{xt}.$$

$$\eta^{xt} = \eta_{xt} + (\eta_{tu} - \xi_{xt})u_x + (\eta_{xu} - \tau_{xt})u_t - \xi_{tu}u_X^2 + (\eta_{uu} - \xi_{xu} - \tau_{tu})u_x u_t - \tau_{xu}u_t^2 - \xi_{uu}u_X^2 u_t$$
$$- \tau_{uu}u_x u_t^2 - \xi_t u_{xx} - \xi_u u_t u_{xx} + (\eta_u - \xi_x - \tau_t)u_{xt} - 2\xi_u u_x u_{xt} - 2\tau_u u_t u_{xt} - \tau_x u_{tt}$$
$$- \tau_u u_x u_{xx},$$

$$\eta^{tt} = \eta_{tt} - \xi_{tt}u_x + (2\eta_{tu} - \tau_{tt})u_t - 2\xi_{tu}u_x u_t + (\eta_{uu} - 2\tau_{tu})u_t^2 - \xi_{uu}u_x u_t^2 - \tau_{uu}u_t^3 - 2\xi_t u_{xt}$$
$$- 2\xi_u u_t u_{xt} + (\eta_u - 2\tau_t)u_{tt} - \xi_u u_x u_{tt} - 3\tau_u u_t u_{tt}.$$

Lie point symmetries are obtained by differentiating the symmetry condition $\Delta(\hat{x}, \hat{t}, \hat{u}, \hat{u}_{\hat{x}}, \hat{u}_{\hat{t}}, \dots) = 0$ when $\Delta(x, t, u, u_x, u_t, \dots) = 0$ wrt ε at $\varepsilon = 0$. We obtain the linearized symmetry condition

$$X\Delta = 0 \text{ when } \Delta = 0.$$

The restriction $\Delta = u_\sigma - \omega(x, t, u, u_x, u_t, \dots) = 0$ lets us eliminate u_σ from $X\Delta = 0$. The we split the remaining terms according to their dependence on derivatives of u to obtain a linear system of ***determining equations*** for ε, τ, and η. The vector space $\mathcal{L}$ of all Lie point symmetry generators of a given PDE is a Lie algebra, although it may not be finite dimensional.

Example 3.4—Consider $u_t = u_x^2$. The linearized symmetry condition is $\eta^t = 2u_x\eta^x$ when $u_t = u_x^2$. Using $u_t = u_x^2$ to eliminate u_t, let's write the linearized symmetry condition explicitly

$$\eta_t - \xi_t u_x + (\eta_u - \tau_t)u_X^2 - \xi_u u_X^3 - \tau_u u_X^4 = 2u_x(\eta_x + (\eta_u - \xi_x)u_x - (\xi_u + \tau_x)u_X^2 - \tau_u u_X^3).$$

After equating the terms that are multiplied by each power of u_x, we are left with the system of determining equations:

$$\tau_u = 0. \quad (7)$$

$$\xi_u + 2\tau_x = 0. \quad (8)$$

$$\eta_u + \tau_t - 2\xi_x = 0. \quad (9)$$

$$\xi_t + 2\eta_x = 0. \quad (10)$$

$$\eta_t = 0. \quad (11)$$

(These equations are ordered with u_x^4 terms first, followed by the u_X^3 terms and so forth. Solving (7) we get

$$\tau = A(x,t),$$

where A is an arbitrary function for the moment. Given this result, it follows that (8) has solution

$$\xi = -2A_x u + B(x,t),$$

so (9) yields

$$\eta = -2A_{xx}u^2 + (2B_x - A_t)u + C(x,t),$$

for some arbitrary functions B and C. When we substitute our results into (10) and (11) we get

$$-4A_{xxx}u^2 + 4(B_{xx} - A_{xt})u + B_t + 2C_x = 0,$$

$$-2A_{xxt}u^2 + (2B_{xt} - A_{tt})u + C_t = 0.$$

Since the functions A, B, and C are independent of u, the above two equations can be decomposed by equating powers of u as follows:

$$C_t = 0,$$

$$B_t + 2C_x = 0,$$

$$2B_{xt} - A_{tt} = 0,$$

$$B_{xx} - A_{xt} = 0,$$

$$A_{xxt} = 0$$

$$A_{xxx} = 0.$$

Using each of the first three of these equations, we get

$$C = \alpha(x), \qquad B = -2\alpha'(x)t + \beta(x), \qquad A = -2\alpha''(x)t^2 + \gamma(x)t + \delta(x),$$

where $\alpha, \beta, \gamma, \delta$ are functions of solely x that are determined by substituting A, B, and C into $B_{xx} - A_{xt} = 0$, $A_{xxt} = 0$, and $A_{xxx} = 0$. Equating equal powers of t and solving the resulting ODEs. We get

$$\xi = -4c_1tx - 2c_2t + c_4\left(\frac{1}{2}x^2 - 2tu\right) + c_6x + c_7 - 4c_8xu - 2c_9u,$$

$$\tau = -4c_1t^2 + c_4xt + c_5t + c_8x^2 + c_9x + c_{10,}$$

$$\eta = c_1x^2 + c_2x + c_3 + c_4xu - c_5u + 2c_6u - 4c_8u^2.$$

Because there are ten arbitrary constants, the Lie algebra is ten dimensional.

Hydon's text outlines a few more examples of this symmetry approach to PDEs in his chapter 8. The method generalizes to PDEs with M dependent variables $u = u(u^1, \dots, u^M)$ and N independent variables $x = x(x^1, \dots, x^N)$. Finally Hydon describes and cites the use of computer algebra tools.

Methods for obtaining exact solutions to PDEs. Just as with ODEs, the symmetry toolset we have acquired for PDEs yields the generators of the PDE. For ODEs, we used these generators, then, to extract invariant solutions. The same applies to PDEs.

In a typical first PDE course one learns to solve PDEs by various methods including: similarity solutions, travelling waves, separation of variables, and so forth involving ansätze. Though not advertised, many of these methods involve nothing more than looking for solutions that are invariant under a particular group of symmetries. For example, PDEs for $u(x,t)$ (like wave equations) whose symmetries include $X_1 = \partial_x$ and $X_2 = \partial_t$ generally have traveling wave solutions of the form $u = F(x - ct)$. These solutions are invariant under the group generated by

$$X = cX_1 + X_2 = c\partial_x + \partial_{t,}$$

because both u and $x - ct$ are invariants. In the same way PDEs with scaling symmetries admit similarity solutions, which are constructed from the invariants of the group.

Let us generalize the idea to any Lie group of symmetries of a given PDE $\Delta = 0$. For now we restrict ourselves to scalar PDEs with two independent variables. Recall that a solution $u = u(x,t)$ is invariant under the group generated by

$$X = \xi\partial_x + \tau\partial_t + \eta\partial_u$$

iff the characteristic vanishes on the solution. That is, every invariant solution satisfies the invariant surface condition

$$Q = \eta - \xi u_x - \tau u_t = 0. \qquad (12)$$

(Review the parallel theory and examples for ODEs.) Usually (12) is very much easier to solve than the original PDE, as we have seen for ODEs. Don't forget that we have to check the solutions we get by plugging them back into the original PDE. For example, the group generated by $X = cX_1 + X_2 = c\partial_x + \partial_t$ has the characteristic

$$Q = -cu_x - u_t.$$

The travelling wave ansatz $u = F(x - ct)$is the general solution of the invariant surface condition $Q = 0$.

Suppose for now that ξ and τ are not both zero. Then the invariant surface condition is a first-order quasilinear PDE that can be solved by the method of characteristics. The characteristic equations are

$$\frac{dx}{\xi} = \frac{dt}{\tau} = \frac{du}{\eta}.$$

If $r(x, t, t)$ and $v(x, t, u)$ are two functionally independent first integrals of the characteristic equation, every invariant of the group is a function of r and v. Often it is convenient to let one invariant play the role of a dependent variable. Suppose WLOG that $v_u \neq 0$. Then the general solution of the invariant surface condition is $v = F(r)$. The solution is now substituted into the PDE $\Delta = 0$ to determine the function F.

If both r and v both depend on u, it is necessary to find out whether the PDE has any solutions of the form $r = c$. These are the only solutions of the invariant surface condition that are not (locally) of the form $v = F(r)$. If r is a function of the independent variables x and t only, then $r = c$ cannot yield a solution $u = u(x, t)$.

Example 3.5—Consider the heat equation $u_t = u_{xx}$. The steps you will see below are highly reminiscent of the kinds of PDEs you will struggle with in graduate electrodynamics (think Jackson's text). It has many symmetries. Consider the two-parameter Lie group of scalings generated by

$$X_1 = x\partial_x + 2t\partial_t, \qquad X_2 = u\partial_u.$$

Every generator of a one-parameter Lie group of scalings is of the form (see 2nd half of chapter 2)

$$X = hX_1 + kX_2,$$

for some constants h and k. Remember that if λ is any nonzero constant, X and λX generate the same one parameter group (see chapter 2). (The group parameter ε is changed, but this does not affect the group.) Therefore if $h \neq 0$ we may assume that $h = 1$ WLOG; if $h = 0$, set $k = 1$.

Suppose that $h = 1$, so that

$$X = x\partial_x + 2t\partial_t + ku\partial_u.$$

The invariant surface condition is

$$Q = ku - xu_x - 2tu_t = 0,$$

which is solved by integrating the characteristic equation

$$\frac{dx}{x} = \frac{dt}{2t} = \frac{du}{ku}.$$

Carrying out the integrations yields $r = xt^{-\frac{1}{2}}$ and $v = ut^{-\frac{k}{2}}$. Because in this case r is independent of u, every invariant solution is of the form $v = F(r)$. This is equivalent to $u = t^{\frac{k}{2}}F\left(xt^{-\frac{1}{2}}\right)$. Then

$$u_t = t^{(k-2)/2}\left(-\frac{1}{2}rF'(r) + \frac{1}{2}kF(r)\right),$$

$$u_{xx} = t^{(k-2)/2}F''(r).$$

Therefore $u = t^{\frac{k}{2}}F\left(xt^{-\frac{1}{2}}\right)$ is a solution to the PDE if

$$F'' + \frac{1}{2}rF' - \frac{1}{2}kF = 0.$$

The general solution of this ODE is

$$F(r) = c_1 U\left(k + \frac{1}{2}, 2^{-\frac{1}{2}}r\right) + c_x V\left(k + \frac{1}{2}, 2^{-\frac{1}{2}}r\right),$$

where $U(p, z)$ and $V(p, z)$ are parabolic cylinder functions. If k is an integer, these functions can be expressed in terms of elementary functions and their integrals. For example, if $k = 0$ then

$$F(r) = c_1 \operatorname{erf}\left(\frac{r}{2}\right) + \mathrm{c}_{2,}$$

where

$$erf(z) = \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-\zeta^2} d\zeta$$

is the error function. If, on the other hand, $k = -1$,

$$F = c_1 e^{-r^2/4} + \mathrm{c}_2 e^{-r^2/4} \int_0^z e^{\zeta^2/4}\, d\zeta.$$

Substituting these results into

$$u = t^{\frac{k}{2}} F\left(xt^{-\frac{1}{2}}\right)$$

results in a large family of solutions which includes the fundamental solution

$$u = t^{-1/2} e^{-x^2/4t}, \qquad k = -1.$$

the error function solution

$$u = erf\left(\frac{x}{2\sqrt{t}}\right), \qquad k = 0,$$

and many other well-known solutions. Get more out of Jackson's "Classical Electrodynamics".

Hydon's chapter nine covers a more PDE examples, including how to get new solutions from known solutions, and the use of nonclassical symmetries. Except for chapter seven, I've rendered Hydon's eleven chapter text more transparent to sophomore students and above. Get it. Do it.

Lagniappe

Once you know the simplest solution to some differential equations, you may employ very clever algebra to extract the rest of the solutions. Enjoy the next two pages and a half. Consider a spring-mass system with spring constant k and mass m. It's total energy is

$$H = \frac{p_x^2}{2m} + \frac{1}{2}m\omega^2 x^2,$$

where the frequency is ν, $\omega = 2\pi\nu$, $\frac{1}{2}m\omega^2 x^2$ is the potential energy, $\frac{p_x^2}{2m} = \frac{m^2 v^2}{2m} = \frac{1}{2}mv^2$ is the kinetic energy, and $H = E$ the total energy. Let's simplify things with

$$P^2 = \frac{p_x^2}{m}, \qquad Q^2 = m\omega^2 x^2.$$

(Q is used in quantum mechanics instead of X, ditto q instead of x.) Then

$$H = \frac{1}{2}(P^2 + Q^2).$$

We turn this (Hamiltonian) H (total energy) into a Schrödinger (differential) equation by letting

$$\hat{P} = \frac{-i\hbar}{\sqrt{m}}\frac{\partial}{\partial q}, \qquad \hat{Q} = \sqrt{m}\omega q.$$

Now we have (a simple quantum model of an electron oscillating from its equilibrium radius.)

$$\hat{H} = \frac{1}{2}(\hat{P}^2 + \hat{Q}^2).$$

Let us study the commutator of $\hat{P}$ and $\hat{Q}$.

$$[\hat{P}, \hat{Q}]\psi = (\hat{P}\hat{Q} - \hat{Q}\hat{P})\psi = \left(-\frac{i\hbar}{\sqrt{m}}\frac{\partial}{\partial q}(\sqrt{m}\omega q) + \sqrt{m}\omega q\frac{i\hbar}{\sqrt{m}}\frac{\partial}{\partial q}\right)\psi = -i\hbar\omega\psi - i\hbar\omega q\frac{\partial\psi}{\partial q} + i\omega q\hbar\frac{\partial\psi}{\partial q}$$

$$= -i\hbar\omega\psi.$$

So $[\hat{P},\hat{Q}] = -i\hbar\omega$. As with electrons orbiting nuclei in discrete orbits with discrete energy levels, the electrons in this model only oscillate at discrete frequencies corresponding to discrete energy levels—the mathematics will bear this out in a page. For each integer n there corresponds an energy E_n (an eigenvalue) and a wave function ψ_n (eigenfunction) which satisfies the differential equation

$$\frac{\hbar^2}{2m}\frac{\partial^2\psi_n}{\partial q^2} - m\omega^2 q^2\psi_n = E\psi_{n,}\ n = 1,2, \ldots.$$

Ladder algebras. For the classical equation $(P^2 + Q^2) = (P + iQ)(P - iQ)$. For the quantum mechanical equation this is so. We instead get

$$(\hat{P} + i\hat{Q})(\hat{P} - i\hat{Q}) = \hat{P}^2 + \hat{Q}^2 + i(\hat{P}\hat{Q} - \hat{Q}\hat{P}) = \hat{P}^2 + \hat{Q}^2 + i(-i\hbar\omega) = \hat{P}^2 + \hat{Q}^2 + \hbar\omega = 2\hat{H} + \hbar\omega\mathbf{I},$$

because $\hat{P}, \hat{Q}$ obviously don't commute (thus being subject to the Heisenberg uncertainty principle). Here $\mathbf{I}$ is the identity operator (doesn't do anything; it's the one, the unit.)

Now consider

$$(\hat{P} - i\hat{Q})(\hat{P} + i\hat{Q}) = \hat{P}^2 + \hat{Q}^2 - i(-i\hbar\omega) = \hat{P}^2 + \hat{Q}^2 - \hbar\omega = 2\hat{H} - \hbar\omega\mathbf{I}.$$

So

$$\hat{P}^2 + \hat{Q}^2 = (\hat{P} \pm i\hat{Q})(\hat{P} \mp i\hat{Q}) \pm \hbar\omega\mathbf{I}.$$

Lastly consider

$$\hat{H}(\hat{P} \pm i\hat{Q})\psi_n = \frac{1}{2}(\hat{P}^2 + \hat{Q}^2)(\hat{P} \pm i\hat{Q})\psi_n = \frac{1}{2}(\hat{P} \pm i\hat{Q})(\hat{P} \mp i\hat{Q})(\hat{P} \pm i\hat{Q})\psi_n \pm \frac{1}{2}\hbar\omega\mathbf{I}(\hat{P} \pm i\hat{Q})\psi_n$$

$$= \frac{1}{2}(\hat{P} \pm i\hat{Q})(\hat{P}^2 + \hat{Q}^2)\psi_n \pm \frac{1}{2}\hbar\omega\mathbf{I}(\hat{P} \pm i\hat{Q})\psi_n = (\hat{P} \pm i\hat{Q})\hat{H}\psi_n \pm \frac{1}{2}\hbar\omega\mathbf{I}(\hat{P} \pm i\hat{Q})\psi_n$$

$$= (\hat{P} \pm i\hat{Q})E_n\psi_n \pm \frac{1}{2}\hbar\omega\mathbf{I}(\hat{P} \pm i\hat{Q})\psi_n.$$

Thus

$$\hat{H}(\hat{P} \pm i\hat{Q})\psi_n = \left(E_n \pm \frac{1}{2}\hbar\omega\right)(\hat{P} \pm i\hat{Q}).$$

So the energy levels of the (harmonic oscillator) electron are equally spaced by $\frac{1}{2}\hbar\omega$.

Let E_0 denote the lowest energy level corresponding to wave function $\psi_{0,}$ the so-called ground state. Then, since we can't get to a lower energy state than the lowest energy state,

$$(\hat{P} - i\hat{Q})\psi_0 = 0.$$

Now,

$$\hat{H}\psi_0 = \frac{1}{2}(\hat{P}^2 + \hat{Q}^2)\psi_0 = E\psi_0$$

implies that

$$\frac{1}{2}\left[(\hat{P} \pm i\hat{Q})(\hat{P} \mp i\hat{Q}) + \hbar\omega\mathbf{I}\right]\psi_0 = E\psi_0,$$

but $(\hat{P} \mp i\hat{Q})\psi_0 = 0$, so

$$\hat{H}\psi_0 = E\psi_0.$$

Therefore $E = E_0 = \frac{1}{2}\hbar\omega$ must be the lowest possible energy state (the ground state). E_0 is also called the zero point energy (energy can never be zero since $\frac{1}{2}\hbar\omega > 0$).

For shorthand, let $\hat{R}_+ = (\hat{P} + i\hat{Q})$, $\hat{R}_- = (\hat{P} - i\hat{Q})$. Then (iterating n times) $\hat{H}\hat{R}_+^n\psi_0 = (E_0 + n\hbar\omega)$. Try plugging in. $\hat{R}_+$ is a raising operator (or creation operator), while $\hat{R}_-$ is a lowering (or annihilation) operator. In quantum physics we have ladder operators for the energy of the harmonic oscillator, angular moment, and so forth, and in particle physics (requiring relativistic quantum mechanics) we have ladder operators to create or destroy particles, *e.g.*, an electron and positron annihilate each other to create a photon. Cool stuff.

If we can find ψ_0 we can find ψ_n purely by our ladder algebra. Let's find that zeroeth eigenfunction.

$$\hat{R}_{-}\psi_0 = (\hat{P} - i\hat{Q})\psi_0 = \left(\frac{d}{dQ} + \frac{1}{\hbar\omega}Q\right)\psi_0 = 0.$$

The solution to this simple ODE is

$$\psi_0 = c_0 e^{-\left(\frac{Q^2}{2\hbar\omega}\right)} H_0(Q) = c_0 e^{-\left(\frac{m\omega q^2}{2\hbar}\right)} H_0(Q),$$

where c_0 is an arbitrary constant and $H_n(Q), n = 0,1.,,,$ are the Hermite polynomials. Then

$$\psi_1 = \hat{R}_{+}\psi_0 = \left(\hat{P} + \frac{1}{\hbar\omega}\hat{Q}\right)\psi_0 = \left(\frac{d}{dQ} - \frac{1}{\hbar\omega}Q\right)H_0(Q)\psi_0 = \left(\frac{d}{dQ} - \frac{Q}{\hbar\omega}\right)c_0 e^{-\left(\frac{Q^2}{2\hbar\omega}\right)} = -\frac{2Q}{\hbar\omega}c_0 e^{-\left(\frac{Q^2}{2\hbar\omega}\right)}.$$

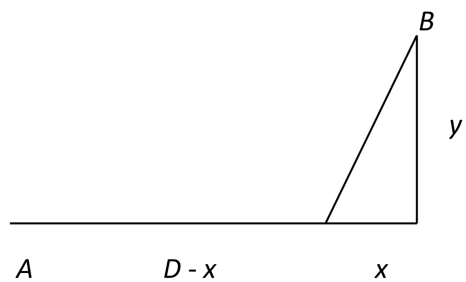
$$\psi_2 = \hat{R}_{+}\psi_1 = -\left(\frac{d}{dQ} - \frac{Q}{\hbar\omega}\right)\frac{2Q}{\hbar\omega}c_0 e^{-\left(\frac{Q^2}{2\hbar\omega}\right)} = -\left(\frac{2}{\hbar\omega} - \frac{2Q}{\hbar\omega}\cdot\frac{Q}{\hbar\omega} - \frac{Q}{\hbar\omega}\cdot\frac{2Q}{\hbar\omega}\right)c_0 e^{-\left(\frac{Q^2}{2\hbar\omega}\right)}$$
$$= \left(\frac{4Q^2}{\hbar^2\omega^2} - \frac{2}{\hbar\omega}\right)c_0 e^{-\left(\frac{Q^2}{2\hbar\omega}\right)}.$$

Indeed $H_0 = 1,\ H_1 = 2Q,\ H_2 = 4Q^2 - 2, \ldots.$ We are generating the Hermite polynomials.

You have now seen that linear algebra, modern (abstract) algebra and differential equations are not disjointed. They are bound together by the unifying methods of symmetry methods in great part because our nature is bound by symmetries. There is more of this unifying still to go. Along the way we will pick up topology and algebraic topology and some differential geometry. You now have at your fingertips powerful tools to treat linear and nonlinear ordinary and partial differential equations and you have deep insights into classical and quantum physics and the problems of mathematical physics.

Part 2. (Accessible to sophomores; required for mathematics)/physics majors up through the postdoctoral research level) Action principles and the calculus of variations.

You, person A is walking along the shore (the $y = 0$ horizontal line) when you notice person B drowning a distance d meters along the shore and y meters into the water. The speed at which you run on sand, v_r is faster than the speed at which you swim, v_s.



You'll run along the line segment $D - x$, and then you'll swim the hypotenuse of the triangle with sides x and y. Treating nature as continuous , there are an infinite number of choices you can make on x before jumping into the water, but there is only one optimal choice, the x that minimizes time. We require only regular calculus to find this point x. The total time to reach B from A is

$$t = \frac{D - x}{v_r} + \frac{\sqrt{x^2 + y^2}}{v_w}.$$

Compute the derivative of t wrt x and set it to 0:

$$\frac{dt}{dx} = -\frac{1}{v_r} + \frac{x}{v_w\sqrt{x^2 + y^2}} = 0.$$

The solution is $x = \left(\left(\frac{v_r}{v_s}\right)^2 - 1\right)^{-\frac{1}{2}} y$. Now hurry! Out of an infinite number of choices we have found the one.

Now tell me what is the shortest distance between two points on the plane? There are infinitely many functions I can imagine connecting the two points. I have asked you to find the function that minimizes distance. This question is in the purview of the calculus of variations. I let one point be the origin $O = (0,0)$ and the other point (x, y). If

$$ds^2 = dx^2 + dy^2 = \left(1 + \left(\frac{dy}{dx}\right)^2\right) dx^2,$$

then the arc-length of the curve $y(x)$ is

$$s = \int_O^{(x,y)} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \int_O^{(x,y)} \sqrt{1 + y'^2} dx = \int_O^{(x,y)} \mathfrak{L}(x, y, y') dx.$$

I'll prove to you that what minimizes s is the solution to the ODE when we derive the equation below

$$\frac{d}{dx}\frac{\partial f}{\partial y'} - \frac{\partial \mathfrak{L}}{\partial y} = 0,$$

where f is a ***functional*** (a function of functions). Let's plug into the above equation.

$$\frac{d}{dx}\frac{\partial f}{\partial y'} = \frac{d}{dx}\frac{\partial}{\partial y'}\sqrt{1 + y'^2} = \frac{d}{dx}\frac{y'}{\sqrt{1 + y'^2}},$$

$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}\sqrt{1 + y'^2} = 0.$$

So

$$\frac{d}{dx}\frac{y'}{\sqrt{1 + y'^2}} = 0.$$

Integration leads to $\frac{y'}{\sqrt{1+y'^2}} = m =$ constant. This is only so if $y' = m$. Then $y(x) = mx + b$.

Action principles are principles that extremize functionals, which are functions of functions, producing either minimal or maximal functions. The fundamental laws of our universe can be expressed in terms of action principles, as well as many emergent laws, like raindrop splashes crowning into symmetric beads and snowflakes forming into six sided shapes. Action principles produce differential equations that are rife with symmetries. This has happened often in particle physics, but one could argue for the reverse case. Nature is rife with symmetries that lead us to differential equations that may be seen as generators of solution curves which extremize some action. We shift our focus a little bit from this "reverse" point of view if we prefer to think of symmetries as measures of invariance, a triangle on a plane being invariant to a handful of rotations and reflections about the plane, spacetime being invariant to continuous Lorentz transformations. This was the path to both special and general relativity. As we have seen for differential equations, invariances can be continuous or discrete. We have an interesting chicken-egg problem. A third point of view is that symmetries thought about in terms of commutators and the concepts of Lie algebras is that symmetries measure interaction. (Anti-) commutators and annihilation and creation operators of particles, be these electrons or strong nuclear force gauge bosons, or theoretical particles such as supersymmetric particles, express interacting particles and fields. A fourth flavor that I see is that commutators measure "entanglement" of variables, space-time being "entangled" in special relativity by Lorentz transformations which in turn "entangle" particles, like the electrons and positrons of the Dirac equation being forced by local relativistic invariance to interact with (gauge boson) photon fields described in terms of vector potentials/covariant derivatives. The mathematicians have now come to see gauge fields as essentially a measure of the curvature of a connection on some fiber bundle, requiring a background in geometry and local and global algebraic topology. At the end of these notes I have an essay/syllabus of good books and ideas that render this paragraph meaningful if it isn't yet so. These notes are themselves a skeleton of the ideas in this paragraph. In any case, action principles are formulated using the calculus of variations.

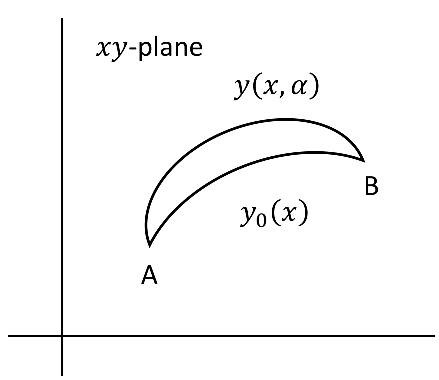
Let's begin developing the calculus of variations. A great, cheap reference book is Elsgolc, now in a Dover edition. Elsgolc develops the theory in parallel with the first year calculus. My approach will be a little more succinct. Consider a function $y(x)$ that is at least twice differentiable in a domain with boundaries x_A and x_B corresponding to points y_A and y_B respectively. We seek to extremize the integral

$$J = \int_{x_A}^{x_B} f(x; y, y')dx.$$

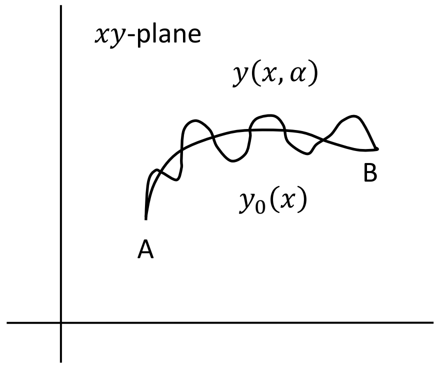
Clearly J is a functional. (Later we shall generalize the theory for $f\big(x; y, y', \dots, y^{(n)}\big).\big)$ Let the solution, whatever it is, be $y_0(x)$. Let $\eta(x)$ any twice differentiable function that vanishes at (x_A, y_A) and (x_B, y_B). That is $\eta(x_A) = \eta(x_B) = 0$. Then the linear combination of $y_0(x)$ and $\eta(x)$,

$$y(x, \alpha) = y_0(x) + \alpha\, \eta(x)$$

satisfies $y(x_A, \alpha) = y_A$ and $y(x_B, \alpha) = y_B$ for all values of the real-valued parameter α. Let's see how this looks like.



The way the figure is drawn, $y(x, \alpha)$ is close to $y_0(x)$ in the sense of $|y(x, \alpha) - y_0(x)|$ is small. In fact, the curves are also close in the sense that $|y'(x, \alpha) - y_0'(x)|$ is small, and so on to higher order derivatives. I might as well generalize the picture. Had I drawn the picture more like



then it is still true that $|y(x,\alpha) - y_0(x)|$ is small, but it is no longer true that $|y'(x,\alpha) - y_0'(x)|$ is small. So let me just define closeness of order k. Two curves $y_0(x)$ and $y_1(x)$ are k-order close if $\left\{y(x,\alpha) - y_0(x), y'(x,\alpha) - y_0'(x), \dots y^{(k)}(x,\alpha) - y_0^{(k)}(x)\right\}$ is small. This definition will be useful we generalize the calculus of variations for functionals depending on $x; y, y', \dots, y^{(k)}$. Getting back to the first picture, the new parametric function $y(x,\alpha)$ represents a continuum of different paths than $y_0(x)$ between the boundary points A and B, one path for each value of α. None of these curves extremizes our functional except the curve for which $\alpha = 0$. Ah! So the functional is really no longer a functional. It is simply a function of the parameter α. That is,

$$J(\alpha) = \int_{x_A}^{x_B} f(x; y(x,\alpha), y'(x,\alpha))dx.$$

We are back to the usual calculus. Let's differentiate J wrt α:

$$\left(\frac{dJ}{d\alpha}\right)_{\alpha=0} = \frac{d}{d\alpha}\int_{x_A}^{x_B} \mathfrak{L}(x; y(x,\alpha), y'(x,\alpha))dx = \int_{x_A}^{x_B} \left\{\frac{\partial f}{\partial y}\frac{\partial y(x,\alpha)}{\partial \alpha} + \frac{\partial f}{\partial y'}\frac{\partial y'(x,\alpha)}{\partial \alpha}\right\} dx = 0.$$

By our construction $\frac{\partial y(x,\alpha)}{\partial \alpha} = \eta(x)$ and $\frac{\partial y'(x,\alpha)}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left[\frac{dy(x,\alpha)}{dx}\right] = \frac{d}{dx}\left[\frac{\partial y(x,\alpha)}{\partial \alpha}\right] = \frac{d\eta(x)}{dx}$. Thus

$$\left(\frac{dJ}{d\alpha}\right)_{\alpha=0} = \int_{x_A}^{x_B} \left[\frac{\partial f}{\partial y}\eta(x) + \frac{\partial f}{\partial y'}\frac{d\eta(x)}{dx}\right] dx = 0.$$

Let's do each integral apart and evaluate the second integral using integration by parts to get

$$\int_{x_A}^{x_B} \frac{\partial f}{\partial y'} \frac{d\eta(x)}{dx} dx = \int_{x_A}^{x_B} \frac{\partial f}{\partial y'} d\eta(x) = \frac{\partial f}{\partial y'} \eta(x) \Big|_{x_A}^{x_B} - \int_{x_A}^{x_B} \eta(x) \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right).$$

The first term is zero at the boundary points. Put the remaining part above with the first term:

$$\int_{x_A}^{x_B} \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right] \eta(x) dx = 0.$$

This is a necessary (not sufficient) condition to have a minimum at $\alpha = 0$. Since $\eta(x)$ is arbitrary,

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0.$$

Voila! We have the Euler-Lagrange differential equation. We have found a condition for finding a path that extremizes J, the Euler-Lagrange differential equation. Physicists normally write $\mathcal{L}$ instead of f, calling $\mathcal{L}$ the Lagrangian, which is the difference of the kinetic energy and potential energy, $\mathcal{L} = T - V$. In physics, **action** is an attribute of the dynamics of a physical system. It is a mathematical functional which takes the trajectory, also called *path* or *history*, of the system as its argument and has a real number as its result. Generally, the action takes different values for different paths. The Euler-Lagrange differential equations extremize the action: $\delta\mathcal{L} = 0$.

Example 1.1—Consider a spring mass system with mass m and spring constant k. The kinetic energy is $T = \frac{1}{2} m\dot{x}^2$ and the potential energy is $V = \frac{1}{2} kx^2$. Thus $\mathcal{L} = \frac{1}{2} m\dot{x}^2 - \frac{1}{2} kx^2$. The Euler-Lagrange differential equation is

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{x}} - \frac{\partial \mathcal{L}}{\partial x} = m\ddot{x} - kx = 0.$$

There is your differential equation of motion. We've seen its quantum mechanical counterpart.

If the Lagrangian does not depend on the independent variable the second order Euler-Lagrange differential equation becomes first order, and as physicists we know that a quantity will be conserved, *e.g.*, angular momentum. To show this consider the following two derivatives: first the total derivative

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

then is the second derivative

$$\frac{d}{dx}\left(y' \frac{\partial f}{\partial y'}\right) = \frac{dy'}{dx} \frac{\partial f}{\partial y'} + y' \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right).$$

Combining these two under the assumption that J doesn't depend on x, ($\frac{df}{dx} = 0$) leads to (with algebra)

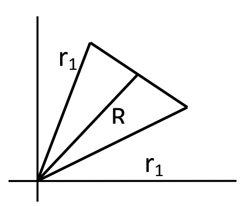
$$\frac{d}{dx}\left(f - y' \frac{\partial f}{\partial y'}\right) = y' \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right].$$

If $y(x)$ minimizes the functional J then the Euler-Lagrange equation becomes

$$\frac{d}{dx}\left(f - y' \frac{\partial f}{\partial y'}\right) = 0, \qquad \text{or} \qquad f - y' \frac{\partial f}{\partial y'} = c,$$

where is an arbitrary constant.

Example 1.2—Consider two objects of mass m_1 and m_2 orbiting each other at r_1 and r_2. The center of mass is R and $r = r_2 - r_1$. The Lagrangian $\mathcal{L} = \frac{1}{2}(m_1 + m_2)\dot{R}^2 + \frac{1}{2}\mu\dot{r}^2 + \frac{1}{2}\mu r^2\dot{\theta}^2 - \frac{G\mu m_2}{r}$, where



$\mu = \frac{(m_1+m_2)}{m_1 m_2}$. Since there is no R in the Lagrangian, then $\frac{d}{dt}\left(\frac{1}{2}(m_1 + m_2)\dot{R}^2\right) = 0$. So the center of mass is an invariant (a symmetry). Ditto for θ. Hence the angular momentum is conserved. That is

$\frac{d}{dt}\left(\frac{1}{2}\mu r^2\dot{\theta}^2\right) = 0$ so $\frac{1}{2}\mu r^2\dot{\theta}^2 = \text{constant} = L$. Here L is the angular momentum of the system. This property of conservation of some quantity was generalized by **Emmy Noether** (generalizing the work of Sophus Lie). Noether's (first) theorem states that any differentiable symmetry of the action of a physical system has a corresponding conservation law. By the way, in this example I snuck in a Lagrangian with more than one independent variable. Can the resulting Euler-Lagrange differential equations for multiple independent variables be coupled? Why yes they can.

Example 1.3—Using two wires of length l, go to your ceiling and suspend to massive bowling balls of mass m from it separated by, say, half a meter. Then couple these bowling balls by attaching a weak spring between them. Let the spring constant be k. Give one ball a nudge towards the other ball, small enough to get the ball swinging but so much as to make it run into the other ball. Eventually, thanks to the spring between the balls, the other ball will begin to swing. Let θ_1 measure the angle of the first ball with respect to its resting position (hanging straight down), and let θ_2 measure the angle of the second ball from its resting position. Then the kinetic energy is

$$T = \frac{1}{2}ml^2\left(\dot{\theta}_1^2 + \dot{\theta}_2^2\right).$$

The potential energy is

$$V = mgl(1 - \cos\theta_1) + mgl(1 - \cos\theta_2) + \frac{1}{2}kl^2(\sin\theta_1 - \sin\theta_2).$$

If these angles are sufficiently small then

$$V \approx mgl(1 - \cos\theta_1) + mgl(1 - \cos\theta_2) + \frac{1}{2}kl^2(\theta_1 - \theta_2).$$

The last term is the source of the coupling. If you change the weak spring into a super long string, then $k = 0$, and the coupling is killed. So $\mathcal{L} = T - V$. The Euler-Lagrange differential equations lead to

$$ml^2\ddot{\theta}_1 + mgl\theta_1 = -2kl^2(\theta_1 - \theta_2),$$

$$ml^2\ddot{\theta}_2 + mgl\theta_2 = 2kl^2(\theta_1 - \theta_2).$$

This folks is a coupled system of differential equations. The sum of the two equations is

$$ml^2\left(\ddot{\theta}_1 + \ddot{\theta}_2\right) + mgl(\theta_1 + \theta_2) = 0.$$

The difference of the top equation minus the second equation is

$$ml^2\left(\ddot{\theta}_1 - \ddot{\theta}_2\right) + mgl(\theta_1 + \theta_2) = -4kl^2(\theta_1 - \theta_2).$$

For kicks, let $\ddot{x} = \left(\ddot{\theta}_1 + \ddot{\theta}_2\right)$ and $\ddot{y} = \left(\ddot{\theta}_1 - \ddot{\theta}_2\right)$. Then

$$ml^2\ddot{x} + mglx = 0,$$

$$ml^2\ddot{y} + mgly = -4kl^2y.$$

Simplifying leads to

$$\ddot{x} + \frac{g}{l}x = 0,$$

$$\ddot{y} + \left(\frac{g}{l} + \frac{4k}{m}\right)y = 0.$$

Let $\omega_x = \sqrt{\frac{g}{l}}$, a frequency, and let $\omega_y = \sqrt{\frac{g}{l} + \frac{4k}{m}} = \sqrt{\frac{g}{l}\left(1 + \frac{4lk}{mg}\right)} \approx= \sqrt{\frac{g}{l}}\left(1 + \frac{4lk}{mg}\right)$ if $4lk \ll mg$. For solutions try

$$x(t) = c_1 \sin \omega_x t + c_2 \cos \omega_x t,$$

$$y(t) = c_3 \sin \omega_y t + c_4 \cos \omega_y t.$$

We specify initial conditions, say, $x(0) = 0$. Thus $c_2 = 0$. For y, specify $y(0) = 0$. Then $c_4 = 0$. Let's say at $t = 0$, the instant you nudge ball 1, it has instantaneous angular speed $\dot{\theta}_1(0) = \dot{\theta}_o$. Then $\dot{x}(0) = \dot{\theta}_o = c_2\omega_x$. So $c_2 = \dot{\theta}_o/\omega_x$. Note if $\dot{\theta}_2(0) = 0$ then $\dot{x}(0) = \dot{y}(0)$. Then $\dot{y}(0) = \dot{\theta}_o = c_4\omega_y$, and so $c_4 = \dot{\theta}_o/\omega_y$. Finally,

$$\theta_1(t) = \theta_o\left(\frac{1}{\omega_x}\sin\omega_x t + \frac{1}{\omega_y}\sin\omega_y t\right),$$

$$\theta_2(t) = \theta_o\left(\frac{1}{\omega_x}\sin\omega_x t - \frac{1}{\omega_y}\sin\omega_y t\right).$$

The motion of one ball affects the motion of the other ball because they are coupled by a weak spring.

Example 1.4—Show and tell. The following is the Einstein-Hilbert action

$$S = \frac{1}{2\kappa}\int R\sqrt{-g}\,d^4x.$$

It leads to the Einstein field equations

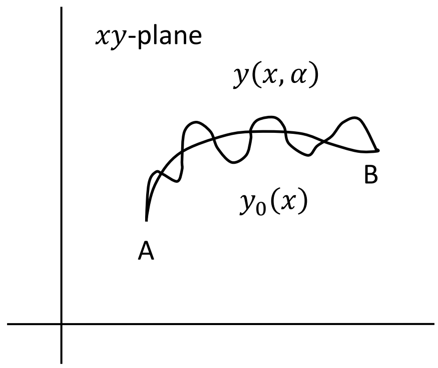
$$R_{\mu\nu} + \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4}T_{\mu\nu}.$$

These really are nothing more than Newton's $\vec{\boldsymbol{F}} = m\vec{\boldsymbol{a}}$, but in a curved four-dimensional curved space-time where the matter/energy fields couple to the space-time curvature (and vice-versa). The key invariance (symmetry) that Einstein found was the indistinguishability of gravitational mass from inertial mass. The Standard Model has its Lagrangian formalism.

If the physical system has constraints, the Euler-Lagrange differential equations need to account for the constraints. (Imaging threading a bead onto a long wire, which we bend into a circle. Suspend the wire circle from the ceiling and spin it. The bead might slide up and down, but always stuck to the wire.) Without proof, if the constraints are specified by $\sum_k a_{ik}\delta q_k = 0$, then the Euler-Lagrange equations become

$$\frac{\partial \mathcal{L}}{\partial q_k} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_k} = \sum_l \lambda_l a_{lk} = 0.$$

There is a way to extend this formulism for problems involving velocity dependent potentials. Finally, if you recall,



the theory extends to

$$J = \int_{x_A}^{x_B} f\big(x; y, y', \dots, y^{(n)}\big)dx.$$

The Euler-Lagrange differential equations become,

$$f_y - \frac{d}{dx}f_{y'} + \frac{d^2}{dx^2}f_{y''} + \cdots + (-1)^n \frac{d^n}{dx^n}f_{y^{(n)}} = 0.$$

The calculus of variations spits out differential equations. Symmetry methods for differential equations, though not fully general, help us unify our treatment of differential equations. Do you feel like you're beginning to get the keys to a large class of universes?

Part 3. (Accessible to possibly juniors, definitely seniors; required for mathematics/physics majors up through the postdoctoral research level) To our differential equations and abstract algebra we know add in topology. (8.5 pages of structure to first unifying example; you'll see the theory is nothing deep.)

Most of the topology we'll see will be point set topology. As with abstract algebra, mathematicians have done a fine job in rendering topology a disjointed witch's brew of meaningless, dry abstraction. Algebraic topology is useful you'll see. These notes are from Gilmore's Dover book.

The origin of point set topology is the real number line. The properties of the real number line have been extended to higher dimensional manifolds, *e.g*. the xy-plane, the surface of the sphere, and even abstracted to build new kinds of "weird" manifolds. Physicist pretty much stick to continuous manifolds. Let's review the properties of the real number line which we will use (an activity you would do in a third or fourth year undergraduate course in real analysis). We will quickly return to Lie groups, Lie algebras and differential equations. Our goal will be to understand Lie's three theorems. The payoff will be deep insight into the nature of quantum fields and general relativity.

Though not necessary, it would be good to review to at least at the level of first year calculus the following concepts: limits, limits of functions, continuous functions, continuity, uniform continuity, sequences, possibly limit infimum and limit supremum, Cauchy sequences, sequences of functions, uniform convergence, and series. A very good and compact source of all of this stuff is chapter 1 (9.5 pages) of "*Real Variables*," Murray R. Spiegel, Schaum's Outline Series. The 9.5 pages of theory is followed by nearly twenty pages of easy, step-by-step examples. Schaum's Outline Series books are cheap, and many of them are damned fine. Here follows the nitty gritty.

A ***neighborhood*** of a point a is the set of all points x such that the distance $\mathrm{d}(x,a) < \delta$, where δ is any given positive number. For the real number line, the distance is defined as $|x-a| < \delta$, If you're on the place we'd have $d(x,y) = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$. If you're on the surface of a sphere you'd have the formula for distance on a sphere, etc. (This concept of distance, norm, or metric, is not stuck to formulas based on the Pythagorean distance. Physicists use L^p norms in quantum physics.)

A ***deleted neighborhood*** of a point a is the set of all points whose distance from a is less than δ, but excluding the point a itself. On the real line this would be $0 < |x - a| < \delta$.

A point $a \in S$ is an ***interior point*** of S if there exists a δ neighborhood of a all of whose points belong to S, *e.g.*, 1 is an interior point of (0,3).

A set is ***open*** if each of its points is an interior point, *e.g.*, (0,3).

If there exists a δ neighborhood of a all of whose points belong to $\tilde{S}$ (the complement of S) then a is called an ***exterior point***, *e.g.*, $\{(-\infty, 0], [3, \infty)\}$ is the complement of (0,3). Any point in the complement of (0,3) is an exterior point.

If every δ neighborhood of a contains at least one point of S and at least one point belonging to $\tilde{S}$ then a is a ***boundary point***. For (0,3) the boundary points are 0 and 3. Every neighborhood of 0 "penetrates" the interior and exterior of (0,3). Ditto 3.

The set of exterior points of S is called the ***exterior*** of S and the set of boundary points of S is called the ***boundary*** of S.

A point $a \in S$ is an ***accumulation point*** or ***limit point*** of S if every deleted δ deleted neighborhood of a contains points of S. Let $S = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$. Every deleted neighborhood of 0 contains points of S, so 0 is a limit point of S, the only limit point in fact.

A set is ***closed*** if it contains all of its limit points, *e.g.*, [0,3] is a closed. The limit points are 0 ,3.

Now we can get to two big theorems we will use.

Theorem 1. The union of any number (possibly infinite) of open sets is open and the intersection of a finite number of open sets is open.

Theorem 2. The union of a finite number of closed sets is closed and the intersection of any number of closed sets is closed.

The proofs of these theorems are easy to find; it's well worth your time. The take away is what these theorems forbid. For the first theorem, you can't have an intersection of an infinite number of open sets result in an open set. For the second theorem, you can't have an intersection of an infinite number of closed sets resulting in a closed set. We have a few more definitions left.

The set consisting of S together with its limit points is the ***closure of*** S, denoted $\bar{S}$.

A set C of open sets (open intervals) is an ***open covering*** of a set S if every point of S belongs to some member of C. If $J \subset S$ is an open covering of S then J is an ***open subcovering*** of S.

(Big one here.) A set S is ***compact*** if every open covering of S has a finite subcovering. For the real line this is equivalent to being closed and bounded, e.g., [0,3]. Ditto for $\mathbb{R}^n$. Getting ahead of ourselves, the continuous group of rotations of the circle $e^{i\theta_1} \cdot e^{i\theta_2} = e^{i\theta_1 + \theta_2}$ is compact on the interval $[0,2\pi]$, but the group of translations ∂_x is noncompact.

We now need to define a topological space and a differentiable manifold, before finally starting to connect to the continuous Lie groups we learned about in Part I. This crap will be pinned down with concrete, accessible examples.

If X is a nonempty set, a class $\mathcal{T}$ of subset of X is a ***topology*** of X iff $\mathcal{T}$ satisfies the following three axioms:

Axiom 1: X and $\mathcal{T}$ belong to $\mathcal{T}$.

Axiom 2: The union of any number (possibly infinite) of set of $\mathcal{T}$ belongs to $\mathcal{T}$.

Axiom 3: The intersection of any two sets of $\mathcal{T}$ belongs to $\mathcal{T}$.

The members of $\mathcal{T}$ are then called $\mathcal{T}$-open sets (or simply open sets), and $(X, \mathcal{T})$ is a ***topological space***. Can you see how this definition of a topology has been ripped off from the real number line? Yes. The real number line is a topology. But so are weird things like $X = \{a, b, s, d, e\}$ with $\mathcal{T}_1 = \{X, \emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d, e\}\}$. $\mathcal{T}_1$ satisfies all of the three axioms, hence $\mathcal{T}_1$ is a topology on X. Now consider $\mathcal{T}_2 = \{X, \emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{b, c, d\}\}$. $\mathcal{T}_2$ is not a topology on X because $\{a, c, d\} \cup \{b, c, d\} \notin \mathcal{T}_2$. $\mathcal{T}_3 = \{X, \emptyset, \{a\}, \{c, d\}, \{a, c, d\}, \{a, b, d, e\}\}$ is not a topology on X because $\{a, c, d\} \cap \{a, b, d, e\} \notin \mathcal{T}_3$.

In this terminology then, an open set S_p containing p is called a neighborhood of p. Symbolically we write $p \in S_p \in \mathcal{T}$.

A ***Hausdorff space***, ***separated space*** or ***T₂*** space is a topological space in which distinct points have disjoint neighborhood's. This is a concept of separation. The real line is a Hausdorff space. Any Euclidean space is a Hausdorff space. For any two distinct point on the real line I can always find tiny enough neighborhoods for each point such that these tiny neighborhoods don't touch each other, *e.g.*, if my two points are 0.000001 and 0.0000011, I can "wrap" 0.000001 in the neighborhood (0.00009,0.00000105) and I can "wrap" 0.0000011 into the neighborhood (0.00000107,42), which are a pair of disjointed neighborhoods.

(Updated, alternative definition of compact space): A space T is ***compact*** if every infinite sequence of points i_i, $i = 1,2,\ldots$ contains a subsequence of points that converges to a point in T. (*This involves the Axiom of Choice; mathematics is screwed because of the Axiom of Choice (see "Mathematics, the loss of certainty," by M. Kline*.))

A ***differentiable manifold*** μ is 1. Hausdorff $(\mathcal{T}, X)$ and 2. A collection Φ of mappings (our Lie symmetry transformations) $\phi_p \in \Phi \ni$ (such that)

μ_1: ϕ_p is a 1-1 mapping of an open set $\mathcal{T}_p$ $(p \in \mathcal{T}_p)$ into an open set in $\mathbb{R}^\eta$.

μ_2: $\cup \mathcal{T}_p = \mathcal{T}$.

μ_3: If $\mathcal{T}_p \cap \mathcal{T}_q \neq \emptyset$, $\phi_p(\mathcal{T}_p \cap \mathcal{T}_q)$ is an open set in $\mathbb{R}^n$ and $\phi_q(\mathcal{T}_p \cap \mathcal{T}_q)$ is an open set in $\mathbb{R}^\eta$ $\ni$ $\phi_q(\mathcal{T}_p \cap \mathcal{T}_q) \neq \phi_p(\mathcal{T}_p \cap \mathcal{T}_q)$.

μ_4: (Maximality) $\phi_p \circ \phi_q^{-1}$ and $\phi_q \circ \phi_p^{-1} \in \Phi$.

The big idea here is to be able to work with the neighborhoods of some potentially ugly or weird space (say the surface of a torus, the surface of a Mobius strip or Klein bottle, $(X, \mathcal{T})$ of our example with the sets being letters, etc.) and map it smoothly onto small neighborhoods in Euclidean space $\mathbb{R}^\eta$ which is much easier to work with. In our case the spaces which we will be mapping to $\mathbb{R}^\eta$ are the spaces made up from η-dimensional continuous Lie groups, *e.g.*, the symmetry group $\mathfrak{so}(3)$ (for some ODE or PDE) being mapped (locally) to $\mathbb{R}^3$. Don't forget this mental picture. To remind ourselves of the tie to differential equations, let me cut and paste from the Hydon text the following definition of a continuous local Lie group.

Suppose that an object occupying a subset of $\mathbb{R}^\eta$, say a solution to an η^{th} order ODE, has an infinite set of symmetries $\Gamma_\varepsilon: \mathrm{x}^{\mathrm{x}} \mapsto \hat{\mathrm{x}}(\mathrm{x}^1, \mathrm{x}^2, \ldots, \mathrm{x}^\eta; \varepsilon)$, $s = 1, \ldots, n$, where ε is a real-valued parameter, and that the following conditions are satisfied:

(L1) Γ_0 is the trivial symmetry, so that $\hat{x}^S = x^S$ when $\varepsilon = 0$. (Identity)

(L2) Γ_ε is a symmetry $\forall$ ε in some neighborhood of zero (a point arbitrarily near zero).

(L3)) $\Gamma_\delta\ \Gamma_\varepsilon = \Gamma_{\delta+\varepsilon}$ ∀ ε sufficiently close to zero. (Closure) (Here the binary operation $\delta * \varepsilon$ means adding the parameters $\delta + \varepsilon$.) (We usually think of these parameters being small, near zero.)

(L4) Each $\hat{x}^S$ may be represented by a Taylor series in ε (in some neighborhood of $\varepsilon = 0$), and therefore $\hat{x}^S(\mathrm{x}^1, \mathrm{x}^2, \dots, \mathrm{x}^\eta; \varepsilon) = x^s + \epsilon\xi^s(\mathrm{x}^1, \mathrm{x}^2, \dots, \mathrm{x}^\eta) + O(\varepsilon^2),\ s = 1, \dots, \eta.$

Then the set of symmetries Γ_ε is a one-parameter, continuous local Lie group. The term "local" refers to the fact that the conditions need only apply in a neighborhood of $\varepsilon = 0$. The term group is used because Γ_ε satisfies the definition of a group.

In slightly more jazzed up language, an η-parameter ***topological group*** (or ***continuous group***) consists of 1-An underlying η-dimensional manifold $\boldsymbol{\mu}$. 2-An operation ϕ mapping each pair of points (β, α) in the manifold into another point in the manifold. 3-In terms of a coordinate system around γ, β, α, we write

$$\gamma^\mu = \phi^\mu(\beta^1, \dots, \beta^\eta, \alpha^1, \dots, \alpha^\eta),$$

where $\mu = 1,2, \dots, \eta$ ∋

$$\phi: \beta \times \alpha \to \gamma = \beta\alpha \text{ is continuous,}$$

$$\psi: \alpha \to \alpha^{-1} \text{ is continuous.}$$

In this definition we are concerned about mapping a pair of elements of a continuous Lie group into a third element of that continuous Lie group by the group operation.

Stop and think about this mental picture. We know from Part I that there are ODEs and PDEs whose Lie symmetries (extracted by the linearized symmetry condition) happen to be the continuous Lie group $\mathfrak{so}(3)$: $X_1 = y\partial_x - x\partial_y,\ X_3 = \frac{1}{2}(1 + x^2 - y^2)\partial_x + xy\partial_y,\ X_3 = xy\partial_x + \frac{1}{2}(-x^2 + y^2)\partial_y$. The commutator algebra of this three-parameter continuous Lie group is the same as the Lie algebra $SO(3)$ defined by the three matrices

$$R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix},\ R_y(\psi) = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix},\ R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which is the symmetry group of the sphere. “Coming from the left” are differential equations leading to the extraction of continuous Lie groups with the same commutator algebra as topological objects like spheres “coming from the right”. We are beginning to see the relatedness of differential equations, abstract algebra and topology and (differential) geometry (*why do we teach these subjects so disjointedly and so dryly with unintuitive abstract mathematical “Psychobabble”? If we must, let abstraction come after a unified foundation*).

Given $SO(3)$, the definition of a continuous group is about doing, for example a continuous rotation by ϕ followed by a continuous rotation by ψ with the result being an element of the group (the sphere is invariant under rotations). It’s generally easier to work with matrix representations of a continuous group than with differential invariants.

If the association between $\mathfrak{so}(3)$ and $SO(3)$ were unique life would be simple, but there are other realizations for the former, and other representations for the latter with the same commutator algebra, *e.g.*, $SU(2)$, the set of 2×2, complex-valued unitary matrices with determinant 1, has the same commutator algebra as $SO(3)$. When the underlying topologies of the continuous groups are the same, we may transform from one representation to another by a similarity transformation. If the global topologies are different, the representations (even though they have the same commutator algebra) are topologically inequivalent. (By the way, changes in global topology cause observable physical effects—see Bohm Aharonov effect.)

Let me give you a flavor of the difference between local and global topology. A topological object with $SU(2)$ symmetry is as different as the sphere with $SO(3)$ as the surface of the sphere is different than the plane generated by translations ∂_x, ∂_y, an issue of global topology. Walk straight along the equator of a sphere and you will return to your initial starting position. Clearly this does not happen if you're on an infinite two-dimensional plane. Up to point set topology, LOCALLY a small patch of the surface of a two-dimensional sphere looks like a little patch of the plane, an issue of. The mathematics of Sophus Lie is local in this sense. The trigonometric functions are linearized in Taylor series: $\sin\theta \approx \theta$and $\cos\theta \approx 1$ when $\theta \ll 1$. A little more structure and we'll get to an example that will feel like an example from Part I.

In a topological (continuous) group, μ we have:

$$\gamma^\mu = \phi^\mu(\beta, \alpha), \qquad \alpha, \beta, \gamma \in \mu \quad \text{(closure)}$$

$$\phi^\mu\big(\gamma, \phi(\beta, \alpha)\big) = \phi^\mu\big(\phi, (\gamma, \beta)\big) \;\; \text{(associativity)}$$

$$\phi^\mu(\varepsilon, \alpha) = \alpha^\mu = \phi^\mu(\alpha, \varepsilon) \;\; \text{(identity)}$$

$$\phi^\mu(\alpha, \alpha^{-1}) = \varepsilon^\mu = \phi^\mu(\alpha^{-1}, \alpha) \quad \text{(inverse)}$$

Part I taught us about continuous groups of Lie transformations. Continuous groups of Lie transformations are special cases of continuous groups of transformations. A ***continuous group of transformations*** is (a) An underlying topological space $\mathcal{T}_\eta$, which is an η-dimensional manifold together with a binary mapping $\phi: \mathcal{T} \times \mathcal{T} \to \mathcal{T}$. (b) A geometric space G_N, which is an N-dimensional manifold, and a mapping $f: \mathcal{T} \times G \to G$ which obey postulate (a'): namely that $\mathcal{T}_\eta, \phi$ obey the postulates of a topological group, and postulate (b'): The function $y^i = f^i(\alpha^1, \ldots, \alpha^\eta; x^1, \ldots, x^{N)}$ is continuous with the following additional properties:

Closure: $\alpha \in \mathcal{T}, x \in G \Rightarrow \alpha x\, x \in G \implies y^i = f^i(\alpha^1, \dots, \alpha^\eta; x^1, \dots, x^{N)} \in G_N,$

Associativity: $\beta(\alpha x) = (\beta \circ \alpha)x\, x \in G \implies f^i(\beta, f(\alpha, x)) = f^i(\phi(\beta, \alpha), x),$

Identity: $\varepsilon x = x \implies f^i(\varepsilon, x) = x^i,$

Inverse: $\alpha^{-1}(\alpha x) = \alpha(\alpha^{-1}x) = (\alpha\alpha^{-1}) = x \implies f^i(\alpha^{-1}, f(\alpha, x)) = f^i(\alpha, f(\alpha^{-1}, x))$

$= f^i(\phi(\alpha, \alpha^{-1}), x) = x^i.$

A continuous group of Lie symmetry transformations is a continuous group of transformations acting on itself ($G \equiv \mathcal{T},\ f \equiv \phi$).

Example 1—Two-parameter group of collinear transformations of the straight line R_1 consisting of

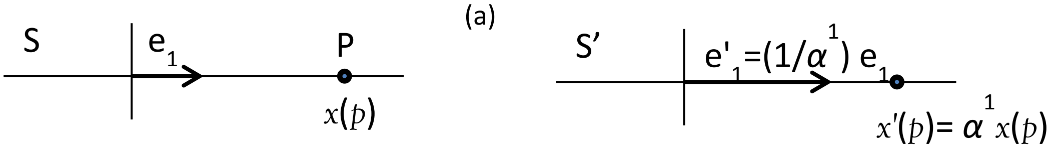
(a) Those operations which change the length of the basis vector.

(b) Those operations which shift the origin.

(c) That operation which changes the orientation of the basis vector (by reflection).

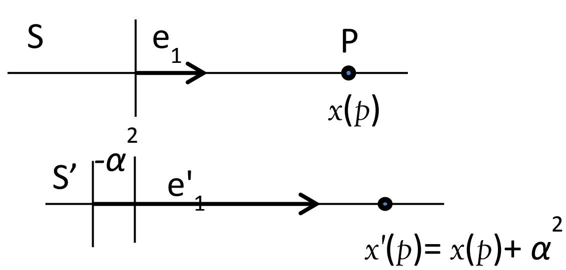
If p is any point in R_1, then wrt to some coordinate system S with basis $\mathbf{e}_1$, it has coordinate $x(p)$. Then under (a), a stretch of the basis $\mathbf{e}_1 \longrightarrow (1/\alpha^1)\mathbf{e}_1$ by $\alpha^1 > 0$, the coordinate p is multiplied by α^1: $x'(p) = \alpha^1 x(p)$. The point P doesn't move; it's invariant. Only its description changes wrt the basis.

(a)



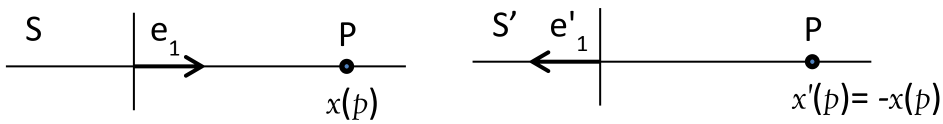
Under (b) a motion of the origin from 0 to $-\alpha^2$, the coordinates of p becomes $x'(p) = x(p) + \alpha^2$.

(b)



Under (c) a reflection of the basis vector $\mathbf{e}_1$, the coordinate of p becomes $x'(p) = -x(p)$.

(c)



Now let's take the pictorial two-parameter group of continuous transformations $\alpha^1 \in \mathbb{R}\backslash 0$, $\alpha^2 \in \mathbb{R}$ and use all of our fancy definitions. Let the operation $\phi(\alpha^1, \alpha^2)$ of the group $\{\alpha^1, \alpha^2\}$ by defined on the coordinate $x(p)$ by $x' = \alpha^1 x + \alpha^2$, $\alpha^1 \neq 0$. The underlying topological space $\mathcal{T}_2$ for this group is the plane $\mathbb{R}^2$ excluding the line$\alpha^1 = 0$ because we defined $\mathbf{e}_1 \longrightarrow (1/\alpha^1)\mathbf{e}_1$. (When we move to the global properties of this continuous group of transformations, the removal of the line $\alpha^1 = 0$ from the plane $\mathbb{R}^2$ will have a deep impact; the space isn't simply connected (not unlike in the Aharonov-Bohm effect)). The geometric space on which our two-parameter group acts on G_1 is our line R_1. <u>I keep saying the "group $\{\alpha^1, \alpha^2\}$, but I have not yet defined the actual rule of the binary operation $\phi(\alpha^1, \alpha^2)$.</u> We now have enough information to determine it. The function f is given by $f(\alpha^1\alpha^2; x) \longrightarrow \alpha^1 x + \alpha^2$. The function ϕ is determined from the group multiplication:

$$x'' = f(\beta^1\beta^2; x), \qquad x' = f(\alpha^1\alpha^2; x)$$

leads to

$$x'' = \beta^1 x' + \beta^2 = \beta^1(\alpha^1 x + \alpha^2) + \beta^2 = f(\beta^1\alpha^1, \beta^1\alpha^2 + \beta^2; x),$$

so,

$$\phi(\beta^1\beta^2, \alpha^1\alpha^2) = (\beta^1\alpha^1, \beta^1\alpha^2 + \beta^2).$$

Now let's check all of the conditions of our fancy definitions for a continuous group. For our example, $\mathcal{T}_2$ is a manifold since each half-plane (the half-plane to the left of $\alpha^1 = 0$, and the half-plane to the right) can be mapped 1-1 onto a part of $\mathbb{R}^2$. From $\phi(\beta^1\beta^2, \alpha^1\alpha^2) = (\beta^1\alpha^1, \beta^1\alpha^2 + \beta^2)$, we identify the coordinates of ϕ

$$\phi^1 = \beta^1\alpha^1,$$

$$\phi^2 = (\beta, \alpha) = \beta^1\alpha^1 + \beta^2.$$

Both ϕ^1 and ϕ^2 are continuous. Let's check:

<u>Closure</u>: $\phi(\beta, \alpha) \in \mathcal{T}_2$ and $\beta^1 \neq 0, \alpha^1 \neq 0 \Rightarrow \phi^1 = \beta^1\alpha^1 \neq 0$. We did this last check because, recall, $\mathcal{T}_2$ excludes the line with first coordinate zero.

<u>Associativity</u>: $\phi\big(\gamma, \phi(\beta, \alpha)\big) = \phi(\gamma^1, \gamma^2; \beta^1\alpha^1 + \beta^2) = (\gamma^1\beta^1\alpha^1, \gamma^1\beta^1\alpha^2 + \gamma^1\beta^2 + \gamma^2) = \phi(\gamma^1\beta^1, \gamma^1\beta^2 + \gamma^2; \alpha^1, \alpha^2) = \phi(\phi(\gamma, \beta), \alpha)$.

Inverse: $\phi\left(\alpha^1, \alpha^2; \frac{1}{\alpha^1}, -\frac{\alpha^2}{\alpha^1}\right) = (1,0) = \phi\left(\frac{1}{\alpha^1}, -\frac{\alpha^2}{\alpha^1}; \alpha^1, \alpha^2\right)$.

We have to similarly check the postulates for a continuous group of transformations. Clearly $G_1 = R_1$ (our line) is a manifold and $f(\alpha^1\alpha^2; x) = \alpha^1 x + \alpha^2$ is continuous in all of its arguments, and f obeys

<u>Closure</u>: $f(\alpha x) \in R_1$.

<u>Associativity</u>: $f(\phi(\beta, \alpha), x) = f(\beta^1\alpha^1 + \beta^2; x) = \beta^1\alpha^1 x + \beta^1\alpha^2 + \beta^2 = f(\beta^1, \beta^2; \alpha^1 x + \alpha^2) = f\big(\beta, f(\alpha, x)\big)$.

<u>Identity</u>: $f(1,0; x) = x$.

Inverse: $f\left(\frac{1}{\alpha^1}, -\frac{\alpha^2}{\alpha^1}; \alpha^1 x + \alpha^2\right) = x.$

We may rewrite all of this structure into matrix representation form with the 1-1 mapping of $\mathcal{T}_2$ into 2× 2 nonsingular matrices

$$(\alpha^1, \alpha^2) \leftrightarrow \begin{pmatrix} \alpha^1 & \alpha^2 \\ 0 & 1 \end{pmatrix}.$$

Then

$$f\colon \begin{pmatrix} x' \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha^1 & \alpha^2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha^1 x + \alpha^2 \\ 1 \end{pmatrix}$$

$$\phi\colon \begin{pmatrix} \gamma^1 & \gamma^2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \beta^1 & \beta^2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha^1 & \alpha^2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \beta^1 \alpha^1 & \beta^2 \\ 0 & 1 \end{pmatrix}.$$

We'll see that matrices will have their advantages. Back in the algebraic form, $x \to x' = \alpha^1 x + \alpha^2$. What happens if we do this twice (square it that is): $(x')^2 = (\alpha^1 x + \alpha^2)^2 = (\alpha^1)^2 x^2 + 2\alpha^1\alpha^2 + (\alpha^2)^2$,

In general,

$$(x')^2 \to (\alpha^1 x + \alpha^2)^N = \sum_{r=0}^{N} \binom{N}{r} (\alpha^1)^r (\alpha^2)^{N-r} x^r.$$

If $N$is a positive integer, the $N + 1$ homogeneous polynomials $(x^N, X^{N-1}, \dots, x, 1)$ can be used as basis for an $(N+1) \times (N+1)$ matrix representations of this *projective* group. These matrix representations are all ***faithful*** (1-1). If N were real, the representations are ∞ dimensional, *e.g.*, here is a faithful 4×4 representation (please refer to Gilmore's Dover text chapter 3).

$$(\alpha^1, \alpha^2) \to \begin{pmatrix} (\alpha^1)^3 & 3(\alpha^1)^2\alpha^2 & 3\alpha^1(\alpha^2)^2 & (\alpha^2)^3 \\ 0 & (\alpha^1)^2 & 2\alpha^1\alpha^2 & (\alpha^2)^2 \\ 0 & 0 & \alpha^1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Quick linear algebra review—This two page review goes well with our picture example. (Think first year physics and/or calculus III.) A ***linear algebra*** is a collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots \in \mathbf{V}$, and a collection $f_1, f_2, \ldots \in \mathrm{F}$, a field together with (α) vector addition, +, (β) scalar multiplication, (τ) vector multiplication $\boxdot \ni$ postulates

[A] $\mathbf{v}_i, \mathbf{v}_j \in \mathbf{V} \Rightarrow \mathbf{v}_i + \mathbf{v}_j \in \mathbf{V}$ (closure under vector addition)

$\mathbf{v}_i + (\mathbf{v}_j + \mathbf{v}_k) = (\mathbf{v}_i + \mathbf{v}_j) + \mathbf{v}_k$ (associativity)

$\mathbf{v}_i + \mathbf{v}_0 = \mathbf{v}_i = \mathbf{v}_0 + \mathbf{v}_i$ (additive identity, $\mathbf{v}_0$)

$\mathbf{v}_i + (-\mathbf{v}_i) = \mathbf{v}_0 = (-\mathbf{v}_i) + \mathbf{v}_i$ (additive inverse)

[B] $f_i \in \mathrm{F}$, $\mathbf{v}_j \in \mathbf{V} \Rightarrow f_i \mathbf{v}_j \in \mathbf{V}$ (closure under scalar multiplication)

$f_i \circ (f_j \circ \mathbf{v}_k) = (f_i f_j)\mathbf{v}_k$ (associativity of scalar multiplication)

$1 \cdot \mathbf{v}_i = \mathbf{v}_i = \mathbf{v}_i \cdot 1$ (scalar multiplicative identity)

$f_i \circ (\mathbf{v}_k + \mathbf{v}_k) = f_i \circ \mathbf{v}_k + f_i \circ \mathbf{v}_i$; $\{f_i + f_j\} \circ \mathbf{v}_k = f_i \circ \mathbf{v}_k + f_j \circ \mathbf{v}_k$ (bilinearity)

<u>STOP HERE</u> and we have the definition of a ***linear vector space***, but let's keep going to [C].

[C] $1, \mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V} \Rightarrow \mathbf{v}_1 \boxdot \mathbf{v}_2 \in \mathbf{V}$

$(\mathbf{v}_1 + \mathbf{v}_2) \boxdot \mathbf{v}_3 = \mathbf{v}_1 \boxdot \mathbf{v}_2 + \mathbf{v}_2 \boxdot \mathbf{v}_3$; $\mathbf{v}_1 \boxdot (\mathbf{v}_2 + \mathbf{v}_3) = \mathbf{v}_1 \boxdot \mathbf{v}_2 + \mathbf{v}_1 \boxdot \mathbf{v}_3$ (bilinearity)

<u>STOP HERE</u> and we have a ***linear algebra***.

We can add additional postulates "[D]" but this will lead to different types of algebras. We can add the postulate that $\mathbf{v}_1 \boxdot \mathbf{v}_2 = \mathbf{v}_2 \boxdot \mathbf{v}_1$ or the postulate that $\mathbf{v}_1 \boxdot \mathbf{v}_2 = -\mathbf{v}_2 \boxdot \mathbf{v}_1$. Clearly these are mutually inconsistent choices. But what would motivate adding one of the above choices over another?

In vector calculus, if we let $\boxdot = \times$, we know the vector cross product: $\mathbf{v}_1 \times \mathbf{v}_2 = -\mathbf{v}_2 \times \mathbf{v}_1$, hence we might chose$\mathbf{v}_1 \boxdot \mathbf{v}_2 = -\mathbf{v}_2 \boxdot \mathbf{v}_1$and be done with it, BUT there are **other** types of important linear algebras. The set of real-valued, $n \times n$ matrices is a linear algebra under matrix addition and scalar multiplication by real numbers; that is, it satisfies postulates [A], [B] and [C]. If we adjoin the usual definition of matrix multiplication by letting $\boxdot$ be defined by $(A \boxdot B)_{ik} = \sum_{j=1}^{n} A_{ij}B_{jk}$ the space becomes an associative algebra, the identity element being $(\mathbf{I})_{ik} = \delta_{ik}$. We're about to run into problems. If we restrict ourselves to the set of $n \times n$ real symmetric matrices obeying $\left(S_{ij}\right)^t$=$S_{ji} = +S_{ij}$, where the t stands for matrix transpose, then we can't use normal matrix multiplication because the product of two symmetric matrices is not generally symmetric. That is, if matrices A, B are $n \times n$ real symmetric matrices, then generally $(A \boxdot B)_{ik} = \sum_{j=1}^{n} A_{ij}B_{jk} \neq$ real symmetric matrix. If we define $\boxdot$ by anticommutation, $A \boxdot B = \{A, B\} = AB + BA$ we get a linear algebra, and $A \boxdot B = BA$. So for these objects we have to choose postulate $\mathbf{v}_1 \boxdot \mathbf{v}_2 = \mathbf{v}_2 \boxdot \mathbf{v}_1$. For objects that are $n \times n$ real <u>*anti*</u>symmetric matrices $A^t = -A$, $A_{ij} = -A_{ji}$, the operation $\boxdot$ must be defined by $[A, B] = AB - BA = -[B, A]$, <u>*our old friend the commutator*</u>. If we add $[A, \beta B + \gamma C] = \beta[A, B] + \gamma[A, C]$, we have a (matrix) Lie algebra! It obeys the (matrix) Jacobi identity $\big[A, [B, C]\big] + \big[C, [A, B]\big] + \big[B, [C, A]\big] = 0$. So the nature of the object determines our choices for the additional postulates [D].

In example 1.4 we looked at a change of basis. In general terms we have a vector $\mathbf{v} = \sum v^i \mathbf{e}_i$ in coordinate system S. Changing coordinate systems to S' we have

$$\mathbf{e}'_j = \sum A^i_j \mathbf{e}_i = A^i_j \mathbf{e}_i.$$

We may drop the summation symbol, the addition being implicit. Notice the matrix A is on the left of the basis vectors. We say basis vectors transform in an ***covariant*** way. On the other hand

$$\mathbf{v} = v^i \mathbf{e}_i = v'^j \mathbf{e}'_j = v'^j A^i_j \mathbf{e}_i = \left(v'^j A^i_j\right) \mathbf{e}_i,$$

where I only added the parenthesis for visual effect. We see that the matrix A is to the right of the vector components v'^j. Thus we say that vectors transform in a contravariant way. Summing up

$$\mathbf{e}' = \boldsymbol{A}\mathbf{e}, \qquad \mathbf{v} = v^i \boldsymbol{A}.$$

We have, effectively, two bases, a "dual basis" in linear algebra, and it helps us define the dot product. Basis vectors with low indices, $\mathbf{e}_i$ transform covariantly, while basis vectors $\mathbf{e^i}$ transform contravariantly. This structure is essential for special and general relativity:

$$\mathbf{a} \cdot \mathbf{b} \equiv (a^0, a^1, a^2, a^3) \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} b^0 \\ b^1 \\ b^2 \\ b^3 \end{pmatrix} = a^0 b^0 - a^1 b^1 - a^2 b^2 - a^3 b^3,$$

where the 4×4 matrix is the metric of special relativity (the Minkowski metric). We bury this structure of the dot product in ordinary, 1st year calculus based physics because the metric tensor is the identity matrix.

There are five more definitions to go in this chapter, but they will have context thanks to our example 1. (Gilmore's Dover text gets into how to discover most of the global properties of a continuous group from its properties near the origin).

A ***local continuous group*** is a manifold $\mathcal{M}$ together with a binary operation ϕ which is defined on certain points $\beta, \alpha \in \mathcal{M}$ (in particular, in a neighborhood of the identity) with the properties

1- $\gamma = \phi(\beta, \alpha) \in \mathcal{M}$ when $\phi(\beta, \alpha)$ is defined, and ϕ is continuous where defined.

2-$\alpha \to \alpha^{-1}$ is continuous when defined.

3-When $\phi(\gamma, \beta)$ and $\phi(\beta, \gamma)$ are defined, $\phi[\phi(\gamma, \beta), \alpha] = \phi[\gamma, \phi(\beta, \alpha)]$ if defined.

4-There is an identity element $\varepsilon \ni \phi(\varepsilon, \alpha) = \phi(\alpha, \varepsilon) = \alpha \in \mathcal{M}$.

5-If α^{-1} is defined, then $\phi(\alpha, \alpha^{-1}) = \phi(\alpha^{-1}, \alpha) = \varepsilon$.

These are the properties of a neighborhood of the identity ε of a continuous group. Warning: Since multiplication and inversion may correspond to points outside the neighborhood of ε, then multiplication and inversion are not defined in such a neighborhood. It is not always true that a local continuous group can be embedded in a global topological group.

Two local continuous groups $(\mathcal{M}, \phi)$ and $(\mathcal{M}', \phi')$ **are locally isomorphic** if (1) there is a 1-1 mapping $\mathcal{M} \to \mathcal{M}'$, (2) If $\alpha, \beta \in \mathcal{M}$ and If $\alpha', \beta' \in \mathcal{M}'$, then $\alpha \circ \beta$ and α^{-1} are defined iff $\alpha' \circ \beta'$ are defined. In addition, the manifold isomorphism must preserve the group operation $(\alpha \circ \beta)' = \alpha' \circ \beta'$.

As I pointed out in a previous discussion, local continuous groups and local Lie groups arise when we linearized the group structure, *e.g.*, the trigonometric terms in the rotation group $SO(3)$ being linearized, $\sin\theta \approx \theta, \cos\theta \approx 1$. We are investigating the properties of the group only near its identity.

We will have occasional use for local compactness. A space is ***locally compact*** if, around and point p a neighborhood can be found whose closure (the neighborhood with its boundary, all its limit points) lies within the set. Our example is locally compact.

Lie groups. Let $\alpha_0 = \varepsilon$ be the identity and let β be some other point in the connected component of a continuous group (the part of the group where any two points can be joined by a line). Then ε and β can be joined by a line lying entirely within the connected component. Choose points

$$\alpha_0 = \varepsilon, \alpha_1, \alpha_2, \dots, \alpha_\infty = \beta$$

on the line with the following properties:

1-α_i and α_{i+1} lie within a common neighborhood.

2-$\alpha_{i+1} \circ \alpha_i^{-1}$ lies inside some neighborhood of the identity ε of the identity ε for each value of i. This can always be done with a countable number of group operations α. Then the group operation β can be written

$$\beta = \alpha_\infty \dots (\alpha_3 \circ \alpha_2^{-1}) \circ (\alpha_2 \circ \alpha_1^{-1})(\alpha_1 \circ \alpha_0^{-1}) \circ \varepsilon.$$

Plainly, β is a product of group operations near the identity. We are interested in studying the function

$$\phi\left(\alpha_{i+1}, \alpha_i^{-1}\right) = \phi(\alpha_i + \delta\alpha, \alpha_i^{-1})$$

with small $\delta -$ think Taylor series expansion. Then we must demand that ϕ be differentiable.

A ***Lie group*** is the connected component of a continuous group in which the composition function ϕ is analytic on its domain of definition. (Hilbert's fifth problem showed that the requirement of analyticity is unnecessarily stringent.) Lie groups of transformations and local Lie groups are similarly defined.

We are now going to start moving fast towards tying Lie's three theorems with the symmetry methods for differential equations we learned about in Part I. The divorce between differential equations and algebraic topology (of which we've only done local algebraic topology) could happen because Lie groups and Lie algebras have a mathematics of their own. Lie's three theorems provide a mechanism for constructing the Lie algebra of any Lie group, instead of us sloppily manhandling all of the possible commutators of all of the symmetry generators. The theorems also characterize the properties of a Lie algebra. More than that, I will show you that this mathematics underlies our general relativity (and other theories of gravity) and quantum fields (the ones we know of and the ones we invent). We will be working with and developing a practical example very quickly this time.

Let $F(p)$ be any function defined on all points $p \in G$. Once we choose a coordinate system in G, we can assign each point p an N-tuple of coordinates:

$$p \text{ (in coordinate system } S) \longrightarrow x^1(p), x^2(p), \dots, x^N(p).$$

The function $F(p)$ can then be written in terms of the parameters $x^i(p)$ in coordinate system S:

$$F(p) = F^S[x^1(p), x^2(p), \dots, x^N(p)].$$

In some other coordinate system S' the coordinates of p will change. Therefore the structural form of the function must change to preserve the fixed value at point p

$$F(p) = F^{S'}\left[{x'}^1(p), {x'}^2(p), \dots, {x'}^N(p)\right]$$

How do we determine $F^{S'}$ if we know F^S? The two coordinate systems S and S' in G are related by an element of the Lie group of transformations

$${x'}^j(p) = f^j[\alpha, x(p)].$$

We already know F^S in terms of $x(p)$. To know $F^{S'}$ in terms of $x'(p)$ we must solve $x(p)$ in terms of $x'(p)$:

$$x(p) = f^i[\alpha^{-1}, x'(p)].$$

Then the complete solution to our problem is

$$F^{S'}\left[{x'}^1(p), {x'}^2(p), \dots, {x'}^N(p)\right] \qquad (1)$$
$$= F^S\left[f^1\left(\alpha^{-1}, x'(p)\right), f^2\left(\alpha^{-1}, x'(p)\right), \dots, f^N\left(\alpha^{-1}, x'(p)\right)\right].$$

This solution is not very useful. Great simplification happens if we stay near the identity. For a Lie group operation $\delta\alpha^\mu$ near the identity 0, the inverse given by $(\delta\alpha^{-1})^\mu = -\delta\alpha^\mu$. We can then write the coordinates $x^j(p)$ as follows

$$x^j(p) = f^j[(\delta\alpha^{-1}), x'(p)] = f^j[-\delta\alpha^\mu, x'(p)] = f^j[0, x'(p)] + \left.\frac{\partial f^j[\beta, x'(p)]}{\partial \beta^\mu}\right|_{\beta=0} (-\delta\alpha^\mu) + \cdots$$

$$\cong x'^j(p) - -\delta\alpha^\mu \left.\frac{\partial f^j[\beta, x'(p)]}{\partial \beta^\mu}\right|_{\beta=0}.$$

When this solution for $x(p)$ in terms of $x'(p)$ is used in (1) we get

$$F^{S'}[x'(p)] = F^S\left[x'^j(p) - \delta\alpha^\mu \left.\frac{\partial f^j[\beta, x'(p)]}{\partial \beta^\mu}\right|_{\beta=0}\right]$$

$$\cong F^S[x'(p)] - \delta\alpha^\mu \left.\frac{\partial f^j[\beta, x'(p)]}{\partial \beta^\mu}\right|_{\beta=0} \frac{\partial}{\partial x'^j} F^S[x'(p)].$$

To lowest order the change in the structural form of F is

$$F^{S'}[x'(p)] - F^S[x'(p)] = \delta\alpha^\mu \left\{- \left.\frac{\partial f^j[\beta, x'(p)]}{\partial \beta^\mu}\right|_{\beta=0} \frac{\partial}{\partial x'^j}\right\} F^S[x'(p)] = \delta\alpha^\mu X_\mu F^S[x'(p)],$$

where

$$X_\mu(x') = - \left.\frac{\partial f^j[\beta, x']}{\partial \beta^\mu}\right|_{\beta=0} \frac{\partial}{\partial x'^j}$$

are the infinitesimal (Lie) generators of infinitesimal displacements of coordinate systems by $\delta\alpha^\mu$, or simply generators. Finite displacements are obtained by repeated applications of the generators.

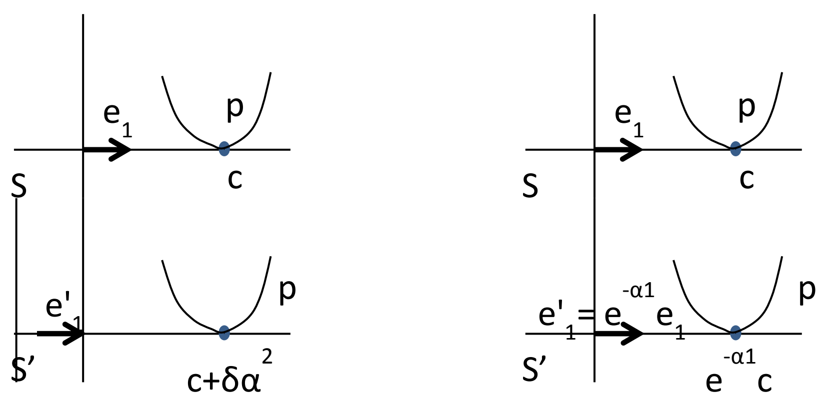
Whoa! In Part I we applied the (prolonged) linearized symmetry condition to differential equations leading to (tangent vector) Lie symmetries from which we then constructed the notion of Lie generators, but even in Part I we noticed the tendency of divorce between differential equations and Lie groups and Lie algebras. We found an ODE and a PDE with symmetry group $SO(3)$. The present approach to Lie generators (which I learned first) seems more fundamental: change coordinates, linearize, get Lie generators/tangent vectors versus get arbitrary differential equations, apply linearized symmetry condition, find tangent vectors/Lie generators. But why should this be so. Maybe because differential equations seem to be pulled out of a giant zoo with no *a priori* relation to any other differential equation, but does any transformation have anything *a priori* to do with any other transformation?

Example 2—Recalling ($f: \mathcal{T} \times G \to G$), from example 1 we had $f^1(\alpha^1, \alpha^2; x) = e^{\alpha^1}x + \alpha^2$. This mapping works on $\mathcal{T} \times G$ no matter what the label of x, *e.g.*, $f^1(\alpha^1, \alpha^2; x') = e^{\alpha^1}x' + \alpha^2$.

$$X_1(x') = -\frac{\partial f(\alpha, x')}{\partial \alpha^1}\bigg|_{\alpha=0} \frac{\partial}{\partial x'} = -x' \frac{\partial}{\partial x'}$$

$$X_2(x') = -\frac{\partial f(\alpha, x')}{\partial \alpha^2}\bigg|_{\alpha=0} \frac{\partial}{\partial x'} = -\frac{\partial}{\partial x'}.$$

Suppose we have a function in S that is $F(S) = (x-c)^2$. Under a displacement of the origin defined by $x' = x + \delta\alpha^2$ the structure of $(x-c)^2$ becomes $\{I + \delta\alpha^2 X_2(x')\}(x'-c)^2 = [x' - (c + \delta\alpha^2)]^2$.



Let's do this in detail for the left hand figures. We have S' is related to S by $x' = x + \delta\alpha^2$. We want to determine the structure of the transformed function. For this simple example, we know it should be

$$F^{S'} = (x' - c - \delta\alpha^2)^2.$$

Let's check this by going through the mathematics. We had

$$F^{S'}[x'(p)] = \left\{\delta\alpha^{\mu}\left\{\mathbf{I} - \frac{\partial f^{j}[\beta, x'(p)]}{\partial\beta^{\mu}}\bigg|_{\beta=0}\frac{\partial}{\partial x'^{j}}\right\}F^{S}[x'(p)]\right\} = \{\mathbf{I} + \delta\alpha^{\mu}X_{\mu}\}F^{S}[x'(p)].$$

Working with the right equation (remembering that we drop powers of $(\delta\alpha^2)^2$ or higher since δ^2 is small)

$$F^{S'}(x') = \{\mathbf{I} + \delta\alpha^2 X_2(x')\}(x'-c)^2 = \left\{\mathbf{I} - \delta\alpha^2\frac{\partial}{\partial x'}\right\}(x'-c)^2 = (x'-c)^2 - 2\delta\alpha^2(x'-c)$$

$$= x'^2 - 2x'c + c^2 - 2\delta\alpha^2 x' + 2\delta\alpha^2 c$$

$$= (x+\delta\alpha^2)^2 - 2(x+\delta\alpha^2)c + c^2 - 2\delta\alpha^2(x+\delta\alpha^2) + 2\delta\alpha^2 c$$

$$= x^2 + \cancel{2\delta\alpha^2 x} + \cancel{(\delta\alpha^2)}^2 - 2cx - \cancel{2c\delta\alpha^2} + c^2 - \cancel{2\delta\alpha^2 x} - \cancel{2(\delta\alpha^2)^2} + \cancel{2c\delta\alpha^2}$$

$$= x^2 - 2cx + c^2 = (x-c)^2 = (x'-c-\delta\alpha^2)^2 = (x'-(c+\delta\alpha^2))^2.$$

Let's now run this work for the right hand figures. We have $x' = xe^{\alpha^1}$.

$$F^{S'}(x') = \{\mathbf{I} + \delta\alpha^1 X_1(x')\}(x'-c)^2 = \left\{\mathbf{I} - \delta\alpha^1 x'\frac{\partial}{\partial x'}\right\}(x'-c)^2 = (x'-c)^2 - 2x'(x'-c)\delta\alpha^1$$

$$= x'^2 - 2cx' + c^2 - 2x'^2\delta\alpha^1 + 2cx'\delta\alpha^1 = x'^2(1-2\delta\alpha^1) - 2cx'(1-\delta\alpha^1) + c^2$$

$$= \left(e^{-\delta\alpha^1}x' - c\right)^2.$$

Is this a stretch? To first order, $e^{-\delta\alpha^1} \sim 1 - \delta\alpha^1$. Then $\left(e^{-\delta\alpha^1}\right)^2 \sim (1-\delta\alpha^1)^2 = 1 - 2\delta\alpha^1 + (\delta\alpha^1)^2 \sim 1 - 2\delta\alpha^1$. Then, going back to the penultimate result, we have

$$F^{S'}(x') = \{\mathbf{I} + \delta\alpha^1 X_1(x')\}(x'-c)^2 = x'^2(1-\delta\alpha^1)^2 - 2cx'(1-\delta\alpha^1) + c^2 = \left(e^{-\delta\alpha^1}x' - c\right)^2.$$

I really can justify this "cheating". Let me show you.

In the first case I got $F^{S'}(x') = (x'-(c+\delta\alpha^2))^2$ with $x' = x + \delta\alpha^2$. If we're interested in finite displacements $x' = x + \alpha^2$, we merely have to exponentiate:

$$F^{S'}(x') = \lim_{N\to\infty} \left\{ \mathbf{I} + \frac{\alpha^2}{N} X_2(x') \right\}^N (x'-c)^2$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} (\alpha^2 X_2(x'))^n (x'-c)^2 = e^{\alpha^2 X_2(x')} (x'-c)^2 = e^{-\alpha^2 \frac{\partial}{\partial x'}} (x'-c)^2$$

$$= \left(1 - \alpha^2 \frac{\partial}{\partial x'}\right)(x'-c)^2 = (x'-c)^2 - 2\alpha^2(x'-c)$$

$$= {x'}^2 - 2x'c + c^2 - 2\alpha^2 x' + 2\alpha^2 c$$

$$= (x+\alpha^2)^2 - 2(x+\alpha^2)c + c^2 - 2\alpha^2(x+\alpha^2) + 2\alpha^2 c$$

$$= x^2 + \cancel{2\alpha^2 x} + (\alpha^2)^2 - 2cx - \cancel{2c\alpha^2} + c^2 - \cancel{2\alpha^2 x} - 2(\alpha^2)^2 + \cancel{2c\alpha^2}$$

$$= x^2 - 2cx + c^2 - (\alpha^2)^2 = (x-c)^2 - (\alpha^2)^2 = \left((x'-c) - \alpha^2\right)^2$$

$$= (x' - (c + \delta\alpha^2))^2.$$

Now for the second case, $F^{S'}(x') = \left(e^{-\delta\alpha^1} x' - c\right)^2$. Let's get the finite displacement.

$$F^{S'}(x') = \lim_{N\to\infty} \left\{ \mathbf{I} + \frac{\alpha^1}{N} X_1(x') \right\}^N (x'-c)^2$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} (\alpha^1 X_1(x'))^n (x'-c)^2 = e^{\alpha^1 X_1(x')} (x'-c)^2 = e^{-\alpha^1 x' \frac{\partial}{\partial x'}} (x'-c)^2$$

$$= e^{-\alpha^1 x' \frac{\partial}{\partial x'}} \left({x'}^2 - 2cx' + c^2\right)$$

$$= \left(1 - \alpha^1 x' \frac{\partial}{\partial x'} + \frac{1}{2}(\alpha^1 x')^2 \frac{\partial^2}{\partial {x'}^2} - \cdots \right)\left({x'}^2 - 2cx' + c^2\right)$$

$$= {x'}^2 - 2cx' + c^2 - 2\alpha^1 {x'}^2 + (\alpha^1 x')^2 + 2c\alpha^1 x'$$

$$= (1 - 2\alpha^1 + (\alpha^1)^2){x'}^2 - 2c(1-\alpha^1)x' + c^2 = (1-\alpha^1)^2 {x'}^2 - 2c(1-\alpha^1)x' + c^2$$

$$= [(1-\alpha^1)x' - c]^2 = \left[e^{-\alpha^1} x' - c\right]^2.$$

I can't do better. This liberty with the Taylor series of the exponential function pervades Lie theory. I'm not 100% comfortable with it. The source author wrote $e^{-\alpha^1 x' \frac{\partial}{\partial x'}}(x'-c)^2 \to \left[e^{-\alpha^1} x' - c\right]^2$.

Infinitesimal generators for a Lie group. A Lie group of transformations acts on itself. With the identification: geometric space $G_N \longrightarrow$ topological space $\mathcal{T}_\eta$, we identify:

$$f(\alpha, x(p)) \longrightarrow \phi(\alpha, \chi(p))$$

$$F[x(p)] = \Phi[\chi(p)].$$

It follows that the infinitesimal generators are

$$X_\mu(x') = -\left.\frac{\partial f^j[\beta, x']}{\partial \beta^\mu}\right|_{\beta=0} \frac{\partial}{\partial x'^j} \longrightarrow X_\mu(\chi') \tag{2}$$

$$= -\left.\frac{\partial \phi^\lambda(\beta, \chi')}{\partial \beta^\mu}\right|_{\beta=0} \frac{\partial}{\partial \chi'^j}.$$

Since a Lie group acting on itself is a nonsingular change of basis,

$$\det \left\| \left.\frac{\partial \phi^\lambda(\beta, \chi'(p))}{\partial \beta^\mu}\right|_{\beta=0} \right\| \neq 0.$$

Example 3—Recall from example 1 that we can replace our group operation with matrix multiplication given the following mapping,

$$(\alpha^1, \alpha^2) \leftrightarrow \begin{pmatrix} \alpha^1 & \alpha^2 \\ 0 & 1 \end{pmatrix}.$$

Any other matrix with the same algebra works. In material that I have yet to present, I have a reason to switch the above matrix for

$$(\alpha^1, \alpha^2) \leftrightarrow \begin{pmatrix} e^{\alpha^1} & \alpha^2 \\ 0 & 1 \end{pmatrix}.$$

This matrix works better for this example. With this new matrix representation we get

$$\phi(\beta^1\beta^2, \chi^1\chi^2) = (\beta^1, \beta^2) \circ (\chi^1, \chi^2) = \begin{pmatrix} e^{\beta^1} & \beta^2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} e^{\chi^1} & \chi^2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{\beta^1+\chi^1} & e^{\beta^1}\chi^2 + \beta^2 \\ 0 & 1 \end{pmatrix}.$$

Then

$$\phi^1(\beta^1\beta^2, \chi^1\chi^2) = \beta^1 + \chi^1,$$

$$\phi^2(\beta^1\beta^2, \chi^1\chi^2) = e^{\beta^1}\chi^2 + \beta^2.$$

(Please go back to example 1 if all of this notation is confusing; it checks. Also keep in mind where you're at—in the continuous Lie group of transformations. Then

$$\begin{pmatrix} \frac{\partial\phi^1}{\partial\beta^1} & \frac{\partial\phi^2}{\partial\beta^1} \\ \frac{\partial\phi^1}{\partial\beta^2} & \frac{\partial\phi^2}{\partial\beta^2} \end{pmatrix}_{\beta=0} = \begin{pmatrix} 1 & \chi^2 \\ 0 & 1 \end{pmatrix}.$$

So, dropping primes, $X_\mu(\chi') = -\frac{\partial\phi^\lambda(\beta,\chi')}{\partial\beta^\mu}\Big|_{\beta=0}\frac{\partial}{\partial\chi'^j}$, together with the matrix above yields

$$-\begin{pmatrix} 1 & \chi^2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \frac{\partial}{\partial\chi^1} \\ \frac{\partial}{\partial\chi^2} \end{pmatrix} = \begin{pmatrix} X_1(\chi) \\ X_2(\chi) \end{pmatrix}.$$

Hence,

$$\chi_1(\chi) = -\frac{\partial}{\partial\chi^1} - \chi^2\frac{\partial}{\partial\chi^2}, \qquad \chi_2(\chi) = -\frac{\partial}{\partial\chi^2}.$$

These are the generators of the continuous Lie group of transformations ($\mathcal{T}$) <u>from the left</u> because $\beta^1\beta^2$ are on the left in

$$\phi(\beta^1\beta^2, \chi^1\chi^2) = (\beta^1, \beta^2) \circ (\chi^1, \chi^2) = \begin{pmatrix} e^{\beta^1} & \beta^2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} e^{\chi^1} & \chi^2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{\beta^1+\chi^1} & e^{\beta^1}\chi^2 + \beta^2 \\ 0 & 1 \end{pmatrix}.$$

How about from the right? ($\chi^1\chi^2$ are on the right of $\phi(\beta^1\beta^2, \chi^1\chi^2)$.) In other words interchange the roles of β^μ with χ^μ in equation (2) resulting in:

$$X_\mu(x') = -\frac{\partial\phi^\lambda(\chi,\chi')}{\partial\chi^\mu}\bigg|_{\chi=0}\frac{\partial}{\partial\beta'^j}.$$

Then,

$$\begin{pmatrix} \dfrac{\partial\phi^1}{\partial\chi^1} & \dfrac{\partial\phi^2}{\partial\chi^1} \\ \dfrac{\partial\phi^1}{\partial\chi^2} & \dfrac{\partial\phi^2}{\partial\chi^2} \end{pmatrix}_{\chi=0} = \begin{pmatrix} 1 & 0 \\ 0 & e^{\beta^1} \end{pmatrix}.$$

Thus from the right (with the X_i as functions of the vector $\beta\{\beta^1.\beta^2\}$),

$$\begin{pmatrix} X_1(\beta) \\ X_2(\beta) \end{pmatrix} = -\begin{pmatrix} 1 & 0 \\ 0 & e^{\beta^1} \end{pmatrix}\begin{pmatrix} \dfrac{\partial}{\partial\beta^1} \\ \dfrac{\partial}{\partial\beta^2} \end{pmatrix} = \begin{pmatrix} -\dfrac{\partial}{\partial\beta^1} \\ -e^{\beta^1}\dfrac{\partial}{\partial\beta^2} \end{pmatrix}.$$

"On the left" we have the X_i as functions of the vector $\chi\{\chi^1,\chi^2\}$ whereas on the right, we have the X_i as functions of the vector $\beta\{\beta^1,\beta^2\}$. Order matters! Especially in physics! Thus there is a "relativistic" thing happening here. The group operation may be interpreted in two different ways. Either (A) the coordinates $\chi(p)$ in coordinates system S are then given in S' which is related back to S by $\chi'(p) = \phi(\alpha, \chi(p))$, OR (**2**), the coordinates $\alpha(p)$ in coordinate system S are given in S' which is related back to S by $\beta'(p) = \phi(\beta(p), \chi)$. In (A) we have the left translation by β. In (B) we have the right translation byχ.

Back in example 2 (with $f: \mathcal{T} \times G \to G$) we derived the generators in the geometric space (G): $X_1(x') = -x'\frac{\partial}{\partial x'}$, $X_2(x') = -\frac{\partial}{\partial x'}$. There was no left or right in

$$X_\mu(x') = -\frac{\partial f^j[\beta, x']}{\partial \beta^\mu}\bigg|_{\beta=0} \frac{\partial}{\partial x'^j}$$

The $x'^j \in G$, while the $\beta^\mu \in \mathcal{T}$. In $\mathcal{T}$, where the Lie group of transformations acts on itself, and we mapped

$$X_\mu(x') = -\frac{\partial f^j[\beta, x']}{\partial \beta^\mu}\bigg|_{\beta=0} \frac{\partial}{\partial x'^j} \longrightarrow X_\mu(\chi') = -\frac{\partial \phi^\lambda(\beta, \chi')}{\partial \beta^\mu}\bigg|_{\beta=0} \frac{\partial}{\partial \chi'^j}.$$

both the β^μ and χ'^j belong to $\mathcal{T}$. Hence in $\mathcal{T}$ we have a left and right translation. Differential equations live in manifolds G. Lie symmetries (topological groups) live in $\mathcal{T}$. Here is the marriage:

$$X_\mu(x') = -\frac{\partial f^j[\beta, x']}{\partial \beta^\mu}\bigg|_{\beta=0} \frac{\partial}{\partial x'^j}.$$

Here is the divorce:

$$X_\mu(x') = -\frac{\partial f^j[\beta, x']}{\partial \beta^\mu}\bigg|_{\beta=0} \frac{\partial}{\partial x'^j} \longrightarrow X_\mu(\chi') = -\frac{\partial \phi^\lambda(\beta, \chi')}{\partial \beta^\mu}\bigg|_{\beta=0} \frac{\partial}{\partial \chi'^j}.$$

Ain't it a shame.

Infinitesimal generators for matrix groups. Way back in Part I, I titillated you with exponentiating a matrix. The generators for matrix groups are defined analogously to how we've defined generators above. If $M(\alpha^1, \alpha^2, \dots, \alpha^\eta)$ is an element of $r \times r$ matrices, we define the infinitesimal generators by

$$X_\mu(r \times r) = \lim_{\alpha^\mu \to 0} \frac{M(0,0,\dots,\alpha^\mu,0,0) - M(0,0,\dots,0)}{\alpha^\mu}.$$

The η generators constructed this way are bases for linear vector spaces since $\lambda^\mu X_\mu(r \times r)$ is also an infinitesimal generator:

$$\lambda^{\mu} X_{\mu}(r \times r) = \lim_{\tau \to 0} \frac{M(\lambda^1 \tau, \lambda^2 \tau, \ldots, \lambda^{\eta} \tau) - M(0,0,\ldots,0)}{\tau}.$$

The generators $X_{\mu}(x)$ and $X_{\mu}(\chi)$ are also bases for linear vector spaces, and the X_{μ} are all linearly independent since each group operation must have a unique inverse.

Example 4—The generators for our 2×2 group of matrices are given by

$$X_1(2 \times 2) = \frac{\partial}{\partial \alpha^1} \begin{pmatrix} e^{\alpha^1} & \alpha^2 \\ 0 & 1 \end{pmatrix}_{\alpha=0} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$X_2(2 \times 2) = \frac{\partial}{\partial \alpha^2} \begin{pmatrix} e^{\alpha^1} & \alpha^2 \\ 0 & 1 \end{pmatrix}_{\alpha=0} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

In this way we may construct the $(N+1) \times (N+1)$ faithful representation of this group with bases $(x^N, X^{N-1}, \ldots, x^2, x^1, x^0)$:

$$X_1 = \text{diagonal}(N, N-1, \ldots, 1, 0),$$

$$X_{2(N+1)\times(N+1)} = \begin{pmatrix} 0 & N-1 & & & & & & & \\ & 0 & & & & & 0 & & \\ & & 0 & & & & & & \\ & & & 0 & & & & & \\ & & & & 0 & & & & \\ & 0 & & & & 0 & 3 & & \\ & & & & & & 0 & 2 & \\ & & & & & & & 0 & 1 \\ & & & & & & & & 0 \end{pmatrix}.$$

Finally (as I showed you explicitly in part I),

$$e^{\alpha^1 X_1(2\times 2)} = \sum_{N=0}^{\infty} \frac{1}{N!} \begin{pmatrix} \alpha^1 & \alpha^2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} = \sum_{N=0}^{\infty} \frac{(\alpha^1)^N}{N!} & \alpha^2 \sum_{N=0}^{\infty} \frac{(\alpha^1)^{N-1}}{N!} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{\alpha^1} & \frac{\alpha^2}{\alpha^1}\left(e^{\alpha^1} - 1\right) \\ 0 & 0 \end{pmatrix}.$$

Yet again we have another parameterization. This one is analytically isomorphic to $\begin{pmatrix} e^{\alpha^1} & \alpha^2 \\ 0 & 1 \end{pmatrix}$ because there is an analytic mapping between them (no proof; see Baker-Campbell-Hausdorff formulas). This leads to another study of Lie groups: how can we classify different Lie groups if even the same group may have more than one analytic structures describing the group multiplication? (Both ϕ and ϕ_1 must have the same local properties. Check them off from the definition and you show equivalence or inequivalence.)

Commutation relations. In this section we peer into the deep underpinnings of general relativity and quantum fields. Recall the commutator: $[A, B] = AB - BA = -[B, A]$, my old friend, the end. Recalling example 2, observe that

$$[X_1(x), X_2(x)] = \left[-x\frac{\partial}{\partial x}, \frac{\partial}{\partial x}\right] = -\frac{\partial}{\partial x} = X_2(x),$$

and "from the left" from example 3,

$$[X_1(\chi), X_2(\chi)] = \left[-\frac{\partial}{\partial \chi^1} - \chi^2\frac{\partial}{\partial \chi^2}, \frac{\partial}{\partial \chi^2}\right] = -\frac{\partial}{\partial \chi^2} = X_2(\chi),$$

and from our matrix group,

$$[X_1(N+1), X_2(N+1)] = X_2(N+1).$$

ALL realizations AND representations have isomorphic commutator relations. There must be a fundamental property of Lie algebras. What is the significance of the commutation relations? I've shown lots of physics cases. Here is the mathematics. If α and β are elements in a commutative (abelian) groups, then

$$\alpha\beta\alpha^{-1} = \beta.$$

If the group is not abelian, γ measures the amount by which $\alpha\beta\alpha^{-1}$ differs from β:

$$\alpha\beta\alpha^{-1} = \gamma\beta,$$

where γ is a group element because

$$\alpha\beta(\beta\alpha)^{-1} = \gamma.$$

The LHS is called the commutator of elements α, β in a group. WOW!!! In mathematics, noncommutativity forces us into the commutator (and the Bianchi identity) and so on). In physics, it distinguishes between classical and quantum physics. The commutator measures the degree of physical "entanglement"; the act of observing one of a pair of noncommuting observables leads to uncertainty in this other observable, *e.g.*, position, momentum and energy, time.

Now enters Sophus Lie. If α and β are near the identity we can expand them in terms of infinitesimal generators (both physics and mathematics):

$$\delta\alpha \to \mathbf{I} + \delta\alpha^{\mu} X_{\mu} + \frac{1}{2}\, \delta\alpha^{\mu} X_{\mu} \delta\alpha^{\nu} X_{\nu} + \cdots,$$

$$\delta\beta \to \mathbf{I} + \delta\beta^{\mu} X_{\mu} + \frac{1}{2}\, \delta\beta X_{\mu} \delta\beta^{\nu} X_{\nu}. + \cdots$$

Now forming the product (keeping things only to first order in $\delta\alpha^{\mu}$ and $\delta\beta^{\mu}$,

$$(\alpha\beta)(\beta\alpha)^{-1} = \left(\mathbf{I} + \delta\alpha^{\mu} X_{\mu} + \frac{1}{2}\, \delta\alpha^{\mu} X_{\mu} \delta\alpha^{\nu} X_{\nu}\right)\Bigg(\mathbf{I} + \delta\beta^{\mu} X_{\mu} \\ + \frac{1}{2}\, \delta\beta X_{\mu} \delta\beta^{\nu} X_{\nu}\Bigg)\left(\left(\mathbf{I} + \delta\beta^{\mu} X_{\mu} + \frac{1}{2}\, \delta\beta X_{\mu} \delta\beta^{\nu} X_{\nu}\right)\left(\mathbf{I} + \delta\alpha^{\mu} X_{\mu} + \frac{1}{2}\, \delta\alpha^{\mu} X_{\mu} \delta\alpha^{\nu} X_{\nu}\right)\right)^{\mathbf{-1}}.$$

Recall that in linearized regime $\left(\mathbf{I} + \delta\alpha^{\mu} X_{\mu} v\right)^{(-1)} = \mathbf{I} - \delta\alpha^{\mu} X_{\mu} v$…. After lots of algebra which you know by now how to do,

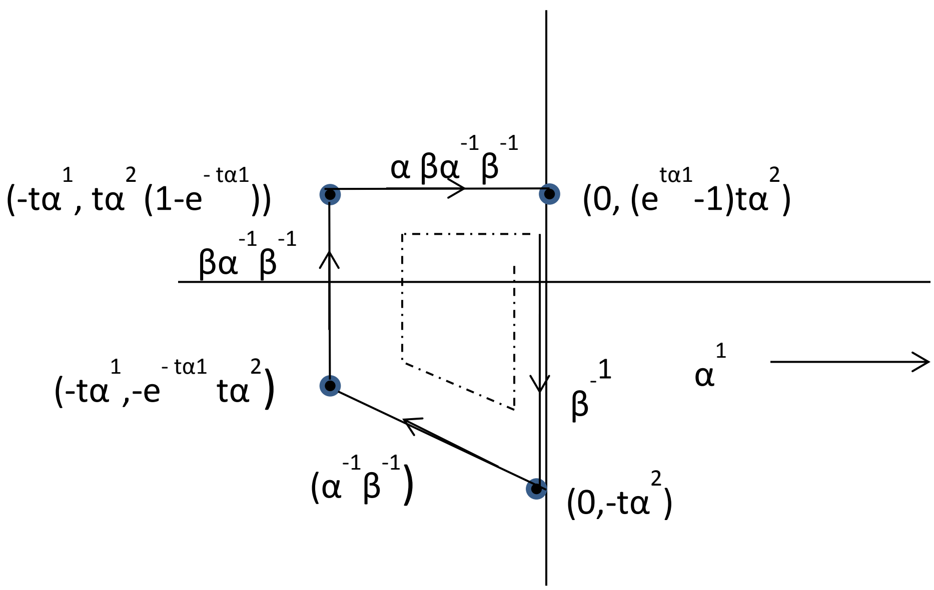
$$(\alpha\beta)(\beta\alpha)^{-1} = \mathbf{I} + \delta\alpha^{\mu}\delta\beta^{\mu}\left(X_{\mu}X_{\nu} - X_{\nu}X_{\mu}\right) = \mathbf{I} + \left[X_{\mu}, X_{\nu}\right].$$

The commutator exists in the vector space of group generators since $(\alpha\beta)(\beta\alpha)^{-1}$ is a member of the group. Therefore the commutator may be expanded in terms of bases X_{λ}:

$$\left[X_{\mu}, X_{\nu}\right] = C_{\mu\nu}^{\lambda} X_{\lambda}.$$

The $C_{\mu\nu}^{\lambda}$ are the ***structure constants***. We've seen use of the structure constants in Part applied to differential equations, namely to the classification of invariant solutions and their corresponding Lie symmetries. I also showed you in the first chapter of Part I (the motivational section) an application of structure constants to "extract" particle spectra. We will do both of these things again and again in latter portions of these notes. The structure constants completely determine the structure of a Lie algebra and almost uniquely determine the structure of the Lie group with that Lie algebra.

Example 5—the BIG PICTURE!



We're inside the continuous (or topological) group. Let $\alpha = (t\alpha^1, 0)$ and $\beta = (0, t\alpha^2)$. Let's take the commutator of α and β. (The outer "loop" has a larger t parameter than the inner "loop".

$$\beta^{-1} = \begin{pmatrix} 1 & -t\alpha^2 \\ 0 & 1 \end{pmatrix} \longrightarrow \alpha^{-1}\beta^{-1} = \begin{pmatrix} e^{-t\alpha^1} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -t\alpha^2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{-t\alpha^1} & -t\alpha^2 e^{-t\alpha^1} \\ 0 & 1 \end{pmatrix}.$$

$$\beta\alpha^{-1}\beta^{-1} = \begin{pmatrix} 1 & t\alpha^2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} e^{-t\alpha^1} & -t\alpha^2 e^{-t\alpha^1} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{-t\alpha^1} & -t^2\alpha^2 e^{-t\alpha^1} + t\alpha^2 \\ 0 & 1 \end{pmatrix}.$$

$$\alpha\beta\alpha^{-1}\beta^{-1} = \begin{pmatrix} e^{t\alpha^1} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} e^{-t\alpha^1} & t^2\alpha^2(1 - e^{-t\alpha^1)} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & t^2\alpha^2\left(e^{t\alpha^1} - 1\right) \\ 0 & 1 \end{pmatrix}.$$

Expand the exponential up to first order.

$$\begin{pmatrix} 1 & \left(e^{t\alpha^1} - 1\right)t\alpha^2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & (1 + (t\alpha^1 - 1)t\alpha^2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & t^2\alpha^1\alpha^2 \\ 0 & 1 \end{pmatrix}.$$

We started at $\beta = (0, t\alpha^2)$ and ended up at $\left(0, t^2\alpha^2\left(e^{t\alpha^1} - 1\right)\right)$. If the algebra were abelian we would have ended up back at where we started, but this algebra is not commutative. Going around the "commutator loop" doesn't get back to where we started by a factor of $\left(e^{t\alpha^1} - 1\right)$.

Differentiate the result wrt t^2.

$$\frac{\partial}{\partial(t^2)} = \alpha(t)\beta(t)\alpha^{-1}(t)\beta^{-1}(t) = \frac{\partial}{\partial(t^2)}\begin{pmatrix} 1 & t^2\alpha^1\alpha^2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \alpha^1\alpha^2 \\ 0 & 1 \end{pmatrix} = \alpha^1\alpha^2 X_2(2 \times 2).$$

<u>Aside on Lie "binding" of classical and quantum physics</u>

(You probably need a bachelor's degree or higher in physics to get the full meaning of this aside. The physics gives context to the senior level mathematician.) We have all of the tools to take a <u>deep</u> look (in a nutshell) of the mathematics binding classical physics to quantum physics and Sophus Lie. We'll do a bonehead example. We build a Lagrangian $\mathcal{L}$. If we are considering our old friend the spring-mass system, we it's $\mathcal{L} = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2$. The principle of least action expressed in variational form (our calculus of variations) leads to the Euler-Lagrange differential equations of motion $\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{x}} - \frac{\partial\mathcal{L}}{\partial x} = m\ddot{x} - kx = 0$. This is an ordinary differential equation. Solutions of this ODE are sine and cosine functions.

The initial conditions let us figure out the arbitrary constants of the general solution of the ODE. There is a whole other way, however, of getting to the solutions than this Lagrangian Mechanics differential equations approach, namely, the Hamiltonian Mechanics approach.

To this end, this is what Hamilton did. He assigned $\frac{\partial \mathcal{L}}{\partial x} = \dot{p}$, where $\dot{p}$ is the time derivative of the momentum p, and substituted $\frac{\partial \mathcal{L}}{\partial x} = \dot{p}$ into the total differential of the Lagrangian $\mathcal{L}(x, p, t)$ to get:

$$d\mathcal{L} = \dot{p}dx + pd\dot{x} + \frac{\partial \mathcal{L}}{\partial t}dt = \dot{p}dx + d(p\dot{x}) - \dot{x}dp + \frac{\partial \mathcal{L}}{\partial t}dt.$$

Rearranging, he got:

$$d(p\dot{x} - \mathcal{L}) = -\dot{p}dx + xdp - \frac{\partial \mathcal{L}}{\partial t}dt.$$

The term in the parenthesis on the LHS is what is called the Hamiltonian, $\mathcal{H}$. That is, $\mathcal{H} = p\dot{x} - \mathcal{L}$. For our spring-mass system our canonical momentum is $p = \frac{\partial \mathcal{L}}{\partial \dot{x}} = m\dot{x}$, where $\dot{x}$ is the speed dx/dt. Note that after rearrangement $\dot{x} = \frac{p}{2m}$. So the Hamiltonian for our spring-mass system is

$$\mathcal{H} = p\dot{x} - \mathcal{L} = p\dot{x} - \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2 = \frac{p^2}{m} - \frac{p^2}{2m} + \frac{1}{2}kx^2 = \frac{p^2}{2m} + \frac{1}{2}kx^2.$$

It happens in this case that $\mathcal{H}$ equals the total energy E, the sum of the kinetic and potential energy. Cool, but where are those solutions I've promised? Almost. We left off with

$$d(p\dot{x} - \mathcal{L}) = -\dot{p}dx + \dot{x}dp - \frac{\partial \mathcal{L}}{\partial t}dt.$$

With $\mathcal{H} = p\dot{x} - \mathcal{L}$, this changes the above equation to

$$d\mathcal{H} = -\dot{p}dx + \dot{x}dp - \frac{\partial \mathcal{L}}{\partial t}dt.$$

By comparing to the definition of the total differential,

$$d\mathcal{H} = \frac{\partial \mathcal{H}}{\partial x}dx + \frac{\partial \mathcal{H}}{\partial p}dp + \frac{\partial \mathcal{H}}{\partial y}dt.$$

Hamilton then made the following identifications (the canonical equations of Hamilton):

$$\frac{\partial \mathcal{H}}{\partial x} = -\dot{p}, \qquad \frac{\partial \mathcal{H}}{\partial p} = \dot{x}, \qquad \frac{\partial \dot{\mathcal{H}}}{\partial t} = -\frac{\partial \mathcal{L}}{\partial t}.$$

Here comes the solution to our spring-mass problem. At time $t_0 = 0$ let the initial position be $x_0 = x(0)$ and the initial momentum be $p_0 = p(0)$. First assume we already know the solution $x(t)$ with Taylor series:

$$x(t) = \sum_{n-0}^{\infty} \frac{1}{n!} x^{(n)}(t)|_{t=t_0} t^n.$$

Now pay attention and note the following pattern for our spring-mass system (then the light bulb will go on!): With (explicitly) $\mathcal{H} = \frac{p^2}{2m} + \frac{1}{2}kx^2$, we have $\frac{\partial \mathcal{H}}{\partial x} = kx$ and $\frac{\partial \mathcal{H}}{\partial p} = \frac{p}{m}$. If the function f is x, (and clearly here x does not explicitly depend on t) the time evolution of x is given by

$$\dot{x} = \frac{d}{dt}x = \{x, \mathcal{H}\}(t = t_0) = \left(\frac{\partial x}{\partial x}\frac{\partial \mathcal{H}}{\partial p} - \frac{\cancel{\partial x}}{\cancel{\partial p}}\frac{\partial \mathcal{H}}{\partial x}\right)_{(t_0)} = \frac{p_0}{m}.$$

(No crap! From freshmen physics $p = m\dot{x} = mv$ and at $t = 0, p_0 = m\dot{x}_0$ Hold your horses.) The time evolution of $\dot{x}$ then is

$$\ddot{x} = \{\dot{x}, \mathcal{H}\}(t_0) = \big\{\{x, \mathcal{H}\}, \mathcal{H}\big\} = \left\{\frac{\partial x}{\partial x}\frac{\partial \mathcal{H}}{\partial p} - \frac{\cancel{\partial x}}{\cancel{\partial p}}\frac{\partial \mathcal{H}}{\partial x} + \frac{\cancel{\partial x}}{\cancel{\partial t}}, \mathcal{H}\right\}_{(t_0)} = \left\{\frac{p}{m}, \mathcal{H}\right\}_{(t_0)} = \frac{1}{m}\{p, \mathcal{H}(t_0)\}$$

$$= \frac{1}{m}\left(\frac{\cancel{\partial p}}{\cancel{\partial x}}\frac{\partial \mathcal{H}}{\partial p} - \frac{\partial p}{\partial p}\frac{\partial \mathcal{H}}{\partial x} + \frac{\cancel{\partial p}}{\cancel{\partial t}}\right)_{(t_0)} = -\frac{1}{m}kx_0$$

(No crap again. We know $m\ddot{x} = ma = kx$. At $t = 0, a_0 = kx_0/m$. Keep going.)

$$\dddot{x} = \{\ddot{x}, \mathcal{H}\}(t_0) = \left\{\{\dot{x}, \mathcal{H}\}, \mathcal{H}\right\}_{(t_0)} = \left\{\left\{\{x, \mathcal{H}\}, \mathcal{H}\right\}, \mathcal{H}\right\}_{(t_0)} = \left\{-\frac{k}{m}x, \mathcal{H}\right\}_{(t_0)} = -\frac{k}{m}\{x, \mathcal{H}\}_{(t_0)} = -\frac{kp_0}{m^2}.$$

Keep going one more time:

$$x^{(4)} = \{\dddot{x}, \mathcal{H}\}(t_0) = \cdots = \left\{-\frac{kp}{m^2}, \mathcal{H}\right\}_{(t_0)} = \frac{k^2 x_0}{m^2}.$$

The iteration is generating the terms $x^{(n)}(t)|_{t=t_0}$. Nice!!! Let's plug in and see if we have enough terms to recognize two distinct Taylor series. The series for $x(t)$ is:

$$x(t) = \frac{1}{0!}x(0) + \frac{p_0}{1!\,m}t - \frac{k}{2!\,m}x_0 t^2 - \frac{k^2}{3!\,m^2}p_0 t^3 + \frac{k^2}{4!\,m^2}x_0 t^4 + \cdots$$

$$= x_0\left(\frac{1}{0!}\cdot 1 - \frac{k}{2!\,m}t^2 + \frac{k^2}{4!\,m^2}t^4\right) + \frac{p_0}{\sqrt{m}}\left(\frac{1}{1!}\left(\frac{k}{m}\right)^{\frac{1}{2}} t - \frac{1}{3!}\left(\frac{k}{m}\right)^{\frac{3}{2}} t^3 + \cdots\right).$$

This is $x(t) = x_0 \cos\omega t + \frac{p_0}{\sqrt{mk}}\sin\omega t$, with $\omega = \sqrt{k/m}$. We've solved $\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{x}} - \frac{\partial \mathcal{L}}{\partial x} = m\ddot{x} - kx = 0$ via Taylor series. We can write:

$$x(t) = x(0) + \frac{p_0}{1!\,m}t - \frac{k}{2!\,m}x_0 t^2 - \frac{k^2}{3!\,m^2}p_0 t^3 + \frac{k^2}{4!\,m^2}x_0 t^4 + \cdots$$

$$= \frac{1}{0!}x(0) + \frac{1}{1!}\{x, \mathcal{H}\}_{(t_0)}t + \frac{1}{2!}\left\{\{x, \mathcal{H}\}, \mathcal{H}\right\}_{t_0} t^2 + \frac{1}{3!}\left\{\left\{\{x, \mathcal{H}\}, \mathcal{H}\right\}, \mathcal{H}\right\}_{(t_0)} t^3 + \cdots$$

$$= \frac{1}{2}x(0)\left(e^{It[\mathcal{H},x]}\Big|_{t=t_0} + e^{-It[\mathcal{H},x]}\Big|_{t=t_0}\right) + \frac{1}{2i}\frac{p_0}{\sqrt{m}}\left(e^{It[\mathcal{H},x]}\Big|_{t=t_0} - e^{-It[\mathcal{H},x]}\Big|_{t=t_0}\right).$$

What? You didn't notice the three faces of the exponential function back in example 2:

$$F^{S'}(x') = \lim_{N\to\infty}\left\{\mathbf{I} + \frac{\alpha^2}{N}X_2(x')\right\}^N (x' - c)^2$$

$$= \sum_{n=0}^{\infty}\frac{1}{n!}\left(\alpha^2 X_2(x')\right)^n (x' - c)^2 = e^{\alpha^2 X_2(x')}(x' - c)^2 = e^{-\alpha^2\frac{\partial}{\partial x'}}(x' - c)^2?$$

In general for some $\mathcal{H}$, what is $\frac{d}{dt}e^{t[\mathcal{H},x]}$?

$$\left(\frac{d}{dt}e^{t[\mathcal{H},x]}\right)_{t-t_0} = \left(\frac{d}{dt}\sum_{n=0}^{\infty}\frac{1}{n!}(\{x,\mathcal{H}\}t)^n\right)_{t-t_0} = \left(\sum_{n=1}^{\infty}\frac{nt^{n-1}[\mathcal{H},x]}{n!}\left(e^{t[\mathcal{H},x]}\right)^{n-1}\right)_{t-t_0}$$

$$= \left([\mathcal{H},x]\sum_{n=1}^{\infty}\frac{1}{(n-1)!}\left(e^{t[\mathcal{H},x]}\right)^{n-1}\right)_{t-t_0}.$$

Let $p = n - 1$. Then

$$\sum_{n=1}^{\infty}\frac{1}{(n-1)!}\left(e^{t[\mathcal{H},x]}\right)^{n-1} \mapsto \sum_{p=0}^{\infty}\frac{1}{p!}\left(e^{t[\mathcal{H},x]}\right)^{p} = e^{[\mathcal{H},x]}.$$

Thus, alas,

$$\left(\frac{d}{dt}e^{t[\mathcal{H},x]}\right)_{t=t_0} = [\mathcal{H},x]_{(t=t_0)}e^{t[\mathcal{H},x]}\Big|_{t=t_0}.$$

This is the classical physics version of the Schrödinger equation.

Turning physics to mathematics, if f is any function of x, p and t, namely $f = f(x,p,t)$, then

$$\frac{d}{dt}f(x,p,t) = \frac{\partial f}{\partial x}\frac{\partial \mathcal{H}}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial \mathcal{H}}{\partial x} + \frac{\partial f}{\partial t} = \{f,\mathcal{H}\} + \frac{\partial f}{\partial t}.$$

where $\{f,H\}$ is the ***Poisson bracket***. If f is any function of x and p, but does not **explicitly** depend on time, t, then $\frac{df}{dt} = 0$ and thus $\frac{\partial f}{\partial t} = -\{f,\mathcal{H}\} = \{\mathcal{H},f\}$. (This was the case for the spring-mass problem, $f \to x$, and thus the function x didn't explicitly depend on t. Yet we *generated* $x(t)$ via "exponentiation".) As we did with the physics example, let's pretend we know $f(t)$. Its Taylor series is

$$f(t) = \sum_{n=0}^{\infty}\frac{f^n(t_0)}{n!}(t-t_0)^n.$$

As with our physics example, now you can check that:

$$f'(t_0) = f^{(1)}(t_0) = \{f, \mathcal{H}\}(t_0),$$

$$f''(t_0) = f^{(2)} = \{f^{(1)}, \mathcal{H}\} = \{\{f, \mathcal{H}\}, \mathcal{H}\}(t_0),$$

$$f'''(t_0) = f^{(3)} = \{f^{(3)}, \mathcal{H}\} = \left\{\{\{f, \mathcal{H}\}, \mathcal{H}\}, \mathcal{H}\right\}(t_0), \ldots,$$

and so on, giving us $f(t) = f(t - t_0)\sum_{n=0}^{\infty}\frac{1}{n!}\{\ldots\{\{f, \mathcal{H}\}, \mathcal{H}\}, \ldots n \text{ times}\}_{t_0}(t - t_0)^n = f(t - t_0)e^{(t-t_0)\{f,\mathcal{H}\}}$. The exponential notation merely represents the series. We have generated $f(t)$ explicitly in t. It follows as before that

$$\left.\frac{df(t)}{dt}\right|_{t-t_0} = \left.\frac{d}{dt}e^{(t-t_0)\{f,\mathcal{H}\}}\right|_{t-t_0} = \left(\{f, \mathcal{H}\}e^{(t-t_0)\{f,\mathcal{H}\}}\right)_{t=t_0}.$$

Now to quantum physics and Lie's work. One historical approach from classical physics to quantum physics was found by assigning classical quantities to differential operators thusly,

$$p_x \longmapsto -i\hbar\frac{\partial}{\partial x} = \hat{p}_x, \;\; x \longmapsto \hat{x} = x, \text{and } \mathcal{H} = E \longmapsto i\hbar\frac{\partial}{\partial t} = \widehat{\mathcal{H}}.$$

The hats denote differential operators. **The Poisson brackets become Lie commutators**.

The classical Hamiltonian of our spring-mass system is

$$\mathcal{H} = E = \frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{p_x^2}{2m} + \frac{1}{2}kx^2$$

Classical physics has been transformed to the quantum physics Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\,\psi(x,t) = \left[\frac{1}{2m}\left(i\hbar\frac{\partial}{\partial x}\right)^2 + \frac{1}{2}kx^2\right]\psi(x,t) = \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(x,t) + \frac{1}{2}kx^2\psi(x,t).$$

As with our classical example, we can do quantum physics via another approach other than solving Schrödinger's equation. To this point in the aside I've been fairly rigorous with the mathematics. I now proceed merely to outline the connection between quantum physics and Lie generators. (To do quantum mechanics you have to be up to speed on the linear algebra of unitary and hermitian matrices and operators.) Given that the Hamiltonian is the infinitesimal (Lie) generator of time translations there is a time evolution operator defined in quantum physics (developed by analogy to classical physics):

$$U(t_0 + dt, t_0) = 1 - \frac{i\widehat{\mathcal{H}}\,dt}{\hbar}$$

Then in textbooks it is shown that

$$U(t_2, t_0) = U(t_2, t_1)U(t_1, t_0), \qquad t_2 > t_1 > t_0.$$

Together these two equations lead to

$$U(t + dt, t_0) = U(t + dt, t)U(t, t_0) = \left(1 - \frac{i\widehat{\mathcal{H}}\,dt}{\hbar}\right)U(t, t_0).$$

Algebra leads to

$$\frac{U(t + dt, t_0) - U(t, t_0)}{dt} = \frac{i\widehat{\mathcal{H}}\,dt}{\hbar}U(t, t_0).$$

In the limit of small dt, we get Schrodinger's equation

$$i\hbar\frac{\partial}{\partial t}U(t, t_0) = \widehat{\mathcal{H}}U(t, t_0).$$

One then goes on in quantum physics considering Hamiltonians which do or do not depend on time.

Did you note that $\frac{d\mathcal{H}}{dt} = 0$ for our classical (and hence quantum) Hamiltonian? In the spring-mass system the energy is a constant. It never changes. It is invariant under transformations of time (the passage of time forward, or run backwards in a movie). The Hamiltonian is the infinitesimal (Lie) generator of time translations. Abstract mathematics, meet physical context.

Historically, three different pictures, all equivalent, of quantum mechanics arose from three different points of view. In the Schrödinger picture the state vectors (wave functions) carry time dependence while the operators stay fixed in time, in the Heisenberg picture is the opposite. In the interaction picture, both state vectors and operators share time dependence. We have seen how this can happen in the study of Lie groups and algebras. Are we dealing with the algebra here $f: \mathcal{T} \times G \to G$ or here G? So some people developed quantum mechanics via Lagrangian mechanics, Hamiltonian mechanics, Hamilton-Jacobi action angle mechanics, Poisson Brackets. Richard Feynman used the Lagrangian in his (the fourth approach) path over histories formulation of quantum physics:

$$\psi(\boldsymbol{x}_2, t_2; \boldsymbol{x}_1, t_1) = \sum_{all\ paths} c e^{-\frac{1}{\hbar}\int_{t_1}^{t_2} L dt}.$$

Lastly, in quantum physics we see the smooth evolution of a smoothly spread wave function pertaining to, say, an electron orbiting an atomic nucleus. The iterated (or "exponentiated" Taylor series) approaches is more reminiscent to particle physics where particles can be created or annihilated (recall the ladder algebra we reviewed in the lagniappe). Each term in the series iterated by a small δt can be seen as particles interacting among each other: an electron being annihilated by a positron, a photon thus being created, or the other way around, the photon splitting into an electron-positron pair.

In quantum field theories, we speak of 2nd quantization. With these notes, up to this point, you already have good unified, symmetry-based foundations in the underlying mathematics and physics. Shall we get back to business? **End Aside**.

Brief aside for the advanced graduate student/postdoc in particles and fields.

Recall the figure of example 5, of taking a round trip in the Lie group space in terms of the commutator, and recall that if the group is nonabelian, the commutator γ measures the amount by which $\alpha\beta\alpha^{-1}$ differs from β:

$$\alpha\beta\alpha^{-1} = \gamma\beta.$$

By taking a round trip in this context by parallel transport, and requiring local gauge invariance, Ryder's text (section 3.6) shows how the covariant derivative arises such that the commutator of the covariant derivative operators leads to gauge fields: (the A_μ are (vector) guage potentials)

$$[D_\mu, D_\nu] = [\partial_\mu - igA_\mu, \partial^\nu - igA_\nu] = -ig\{\partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu]\} = -igG_{\mu\nu}.$$

Then he shows how this mathematics parallels the mathematics of general relativity, the quantity being analogous to the field tensor $G_{\mu\nu}$, the Riemann-Christoffel curvature tensor

$$R^\kappa_{\lambda\mu\nu} = \partial_\nu\Gamma^\kappa_{\lambda\mu} - \partial_\mu\Gamma^\kappa_{\lambda\nu} + \Gamma^\rho_{\lambda\mu}\Gamma^\kappa_{\rho\nu} - \Gamma^\rho_{\lambda\nu}\Gamma^\kappa_{\rho\mu}.$$

I love step-by-step books. "Quantum Field Theory," 2nd ed., L. H. Ryder, Cambridge, 1996 is a step-by-step introduction to QFT. I hope that these notes enrich your understanding of quantum fields and general relativity. Later, after I've covered algebraic topology dealing with global properties of symmetry groups, I'll have an aside on how mathematicians have generalized the concept of vector potential/covariant derivative with differential forms and connections on principle fiber bundles.

Lie's first theorem. (Review examples 1 -3. Keep $f: \mathcal{T} \times G \to G$ and $G \times G$ in mind; the Greek letters are elements of the continuous group $\mathcal{T}$ (η-dimensional) they are the Lie symmetries for us. The coordinates $x'^i(p)$ are for the points p in the continuous geometric group G (N-dimensional). The coordinates system is relative to frame S.

If

$$x'^i(p) = f^i[\alpha^\mu . x(p)], \qquad \mu = 1,2,\dots,\eta, \qquad i = 1,2,\dots,N$$

is analytic (smooth; expandable in a Laurent (or Taylor) series) then

$$\frac{\partial x'^i}{\partial \alpha^\lambda} = \frac{\partial f^i[\alpha, x]}{\partial \alpha^\lambda} = \sum_{\kappa=1}^{\eta} \Psi_\lambda^\kappa(\alpha) u_\kappa^i(x')$$

where $u_\kappa^i(x')$ is analytic.

Proof. If we transform from coordinate system $S \overset{\alpha}{\longrightarrow} S' \overset{\beta}{\longrightarrow} S''$, the coordinates of any point p will change successively from $x'(p)$ to $x'^i(p)$ to $x''^i(p)$. Suppose β is an operation near the identity, $\beta = \delta\alpha$. Then $x''^i(p)$ will be close to $x'^i(p)$, and the difference

$$dx''^i(p) = x''^i(p) - x'^i(p)$$

can be computed in two different ways:

$$x'^i(p) = (x + dx)^i = f^i[\alpha, x],$$

$$x''^i(p) = (x' + dx')^i = f^i[\delta\alpha, x'].$$

The middle terms are differential translations in G. The right hand terms are group actions from $f: \mathcal{T} \times G \rightarrow G$. In terms of f, and in terms of the linearized Taylor translations, the equation for $x'^i(p)$ leads to

$$dx'^i = \left. \frac{\partial f^i[\beta, x']}{\partial \beta^\lambda} \right|_{\beta=0} \delta\alpha^\lambda = \delta\alpha^\lambda u_\lambda^i(x').$$

Clearly we've identified the $u_\lambda^i(x')$ with the $\left. \frac{\partial f^i[\beta, x']}{\partial \beta^\lambda} \right|_{\beta=0}$. Back in $\mathcal{T}$,

$$(\alpha + \delta\alpha)^{\mu} = \phi^{\mu}(\delta\alpha, \alpha)$$

$$d\alpha^{\mu} = \left.\frac{\partial\phi^{\mu}(\beta,\alpha)}{\partial\beta^{\lambda}}\right|_{\beta=0} \delta\alpha^{\lambda} = \delta\alpha^{\lambda}\Theta_{\lambda}^{\mu}(\alpha),$$

where $\Theta_{\lambda}^{\mu}(\alpha)$ is a $\eta \times \eta$ nonsingular matrix identified with $\left.\frac{\partial\phi^{\mu}(\beta,\alpha)}{\partial\beta^{\lambda}}\right|_{\beta=0}$. The inverse of $\Theta_{\lambda}^{\mu}(\alpha)$ is $\Psi(\alpha)$.

$$\Theta_{\lambda}^{\mu}(\alpha)\Psi(\alpha) = \mathbb{I} = \Psi(\alpha)\Theta_{\lambda}^{\mu}(\alpha).$$

Therefore,

$$\delta\alpha^{\mu}\Theta_{\lambda}^{\mu}(\alpha)\Psi_{\mu}^{\lambda}(\alpha) = d\alpha^{\mu}\Psi_{\mu}^{\lambda}(\alpha).$$

This simplifies to

$$\delta\alpha^{\mu} = d\alpha^{\mu}\Psi_{\mu}^{\lambda}(\alpha).$$

In terms of the displacement $d\alpha$near α induced by infinitesimal displacements $\delta\alpha$ at the identity, the displacement dx' at $x'(p)$ is given by. (The result for $\delta\alpha^{\mu}$ is substituted into the equation for dx'^{i}.)

$$dx'^{i} = \delta\alpha^{\lambda}u_{\lambda}^{i}(x') = d\alpha^{\mu}\Psi_{\mu}^{\lambda}(\alpha)u_{\lambda}^{i}(x').$$

Differentiating wrt to α^{μ} results in:

$$\frac{\partial x'^{i}}{\partial\alpha^{\mu}} = \sum_{\lambda=1}^{\mu} \Psi_{\mu}^{\lambda}(\alpha)u_{\lambda}^{i}(x').$$

Since $f^{i}(\alpha,\mu)$ and $\phi^{\mu}(\beta,\alpha)$ are analytic,

$$u_{\lambda}^{i}(x') = \left.\frac{\partial f^{i}[\beta, x']}{\partial\beta^{\lambda}}\right|_{\beta=0},$$

$$\Theta^{\mu}_{\lambda}(\alpha) = \left.\frac{\partial \phi^{\mu}(\beta,\alpha)}{\partial \beta^{\lambda}}\right|_{\beta=0}$$

are also analytic.

Example 6—(The theorem made concrete.) Sticking with examples 1-4, if $\alpha(\alpha^1, \alpha^2)$ is a transformation, and $\delta\alpha = (\delta\alpha^1, \delta\alpha^2)$ is near the identity, then

$$x' = e^{\alpha^1}x + \alpha^2$$

$$x' + dx' = e^{\delta\alpha^1}x + \delta\alpha^2 \approx (1 + \delta\alpha^1)x' + \delta\alpha^2$$

$$x' + dx' = x' + \delta\alpha^1 x' + \delta\alpha^2$$

$$dx' = \delta\alpha^1 x' + \delta\alpha^2$$

Since $\delta\alpha \circ \alpha$ is near the α, we write $(\alpha + d\alpha)^{\mu} = \phi^{\mu}(\delta\alpha^1, \delta\alpha^2; \alpha^1, \alpha^2)$. Recall from our previous examples that $\phi^1 = \alpha^1 + \delta\alpha^1$, and $\phi^2 = e^{\delta\alpha^1}\alpha^2 + \delta\alpha^2$. So

$$\phi^1 = \alpha^1 + \delta\alpha^1 = (\alpha + d\alpha)^1 \implies d\alpha^1 = \delta\,\alpha^1$$

$$\phi^2 = e^{\delta\alpha^1}\alpha^2 + \delta\alpha^2 \approx (1 + \delta\alpha^1)\alpha^2 + \delta\alpha^2 = (\alpha + d\alpha)^2 \Longrightarrow d\alpha^2 = \alpha^2\delta\alpha^1 + \delta\alpha^2.$$

In matrix form,

$$\begin{pmatrix} d\alpha^1 \\ d\alpha^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \alpha^2 & 1 \end{pmatrix} \begin{pmatrix} \delta\alpha^1 \\ \delta\alpha^2 \end{pmatrix}.$$

Inverting and solving for $\delta\alpha^1, \delta\alpha^2$ we get:

$$\begin{pmatrix} \delta\alpha^1 \\ \delta\alpha^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \alpha^2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} d\alpha^1 \\ d\alpha^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\alpha^2 & 1 \end{pmatrix} \begin{pmatrix} d\alpha^1 \\ d\alpha^2 \end{pmatrix}.$$

Then $dx' = \delta\alpha^1 x' + \delta\alpha^2$ which is in G (Recall: $f: \mathcal{T} \times G \to G$) and the results of our solution for $\delta\alpha^1, \delta\alpha^2$ which live in $\mathcal{T}$ as do $d\alpha^1, d\alpha^2$ leads to

$$dx' = x' d\alpha^1 + (-\alpha^2 d\alpha^1 + d\alpha^2).$$

Then

$$\frac{\partial x'}{\partial \alpha^1} = x'^1 - \alpha^2,$$

$$\frac{\partial x'}{\partial \alpha^2} = 1.$$

The formal expression

$$\frac{\partial x'^i}{\partial \alpha^\mu} = \sum_{\lambda=1}^{\mu} \Psi_\mu^\lambda(\alpha) u_\lambda^i(x')$$

can also be used to determine these differential equations:

$$\Theta_\lambda^\mu(\alpha) = \left.\frac{\partial \phi^\mu(\beta,\alpha)}{\partial \beta^\lambda}\right|_{\beta=0} = \begin{pmatrix} \frac{\partial \phi^1}{\partial \beta^1} & \frac{\partial \phi^2}{\partial \beta^1} \\ \frac{\partial \phi^1}{\partial \beta^2} & \frac{\partial \phi^2}{\partial \beta^2} \end{pmatrix}_{\beta=0} = \begin{pmatrix} 1 & \alpha^2 \\ 0 & 1 \end{pmatrix},$$

$$u_\lambda^i(x') = \left.\frac{\partial f^i[\beta, x']}{\partial \beta^\lambda}\right|_{\beta=0} = \begin{pmatrix} \frac{\partial f}{\partial \beta^1} \\ \frac{\partial f}{\partial \beta^2} \end{pmatrix}_{\beta=0} = \begin{pmatrix} x' \\ 1 \end{pmatrix}.$$

Thus we have

$$\begin{pmatrix} \frac{\partial x'}{\partial \alpha^1} \\ \frac{\partial x'}{\partial \alpha^2} \end{pmatrix} = \begin{pmatrix} 1 & -\alpha^2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ 1 \end{pmatrix} = \begin{pmatrix} x' - \alpha^2 \\ 1 \end{pmatrix}.$$

Sophus Lie decomposed partial differential equations into the product of two matrices, one depending on the transformation parameters α^μ and the other depends on the initial conditions $x'^i(p)$. This is a generalization of the problem of finding solutions for systems of simultaneous linear partial differential equations with constant coefficients

$$\frac{\partial x^i}{\partial \tau} = A^i_j x^j(\tau),$$

which can be treated by standard algebraic (Part I) techniques.

Lie's second theorem. The structure constants are constants. If X_μ are generators of a Lie group, then the coefficients $C^\lambda_{\mu\nu}$ given by the commutators $[X_\mu, X_\nu] = C^\lambda_{\mu\nu} X\_\lambda$ are constants. Before proceeding to the proof, note this critical fact: The differential equations

$$\frac{\partial x'^i}{\partial \alpha^\mu} = \sum_{\lambda=1}^{\mu} \Psi^\lambda_\mu(\alpha) u^i_\lambda(x')$$

do not always have a solution. The equations

$$\frac{\partial \psi}{\partial s} = t, \qquad \frac{\partial \psi}{\partial t} = 0$$

have not solutions for

$$\frac{\partial}{\partial t}\frac{\partial \psi}{\partial s} = 1 \neq 0 = \frac{\partial}{\partial s}\frac{\partial \psi}{\partial t}.$$

Note also that the necessary and sufficient conditions for the existence of a unique solution with initial conditions

$$x'^i = f^i[\alpha, x]|_{\alpha=0} = x^i$$

is that all mixed derivatives be equal

$$\frac{\partial^2 x'^i}{\partial\alpha^\mu\partial\alpha^\nu} = \frac{\partial^2 x'^i}{\partial\alpha^\nu\partial\alpha^\mu}.$$

These are called the ***integrability conditions***.

Proof. We apply the integrability conditions to the equation proven in Lie's first theorem

$$\frac{\partial x'^i}{\partial\alpha^\lambda} = \frac{\partial f^i[\alpha,x]}{\partial\alpha^\lambda} = \sum_{\kappa=1}^{\eta} \Psi_\lambda^\kappa(\alpha)u_\kappa^i(x')$$

$$\frac{\partial}{\partial\alpha^\mu}\{\Psi_\nu^\kappa(\alpha)u_\kappa^i(x')\} = \frac{\partial}{\partial\alpha^\nu}\{\Psi_\mu^\lambda(\alpha)u_\lambda^i(x')\}$$

This leads to

$$\frac{\partial\Psi_\nu^\kappa(\alpha)}{\partial\alpha^\mu}u_\kappa^i(x') + \Psi_\nu^\kappa(\alpha)\frac{\partial u_\kappa^i(x')}{\partial\alpha^\mu} = \frac{\partial\Psi_\mu^\lambda(\alpha)}{\partial\alpha^\nu}u_\lambda^i(x') + \Psi_\mu^\lambda(\alpha)\frac{\partial u_\lambda^i(x')}{\partial\alpha^\nu}.$$

Rearrange this result to

$$\Psi_\nu^\kappa(\alpha)\frac{\partial u_\kappa^i(x')}{\partial\alpha^\mu} - \Psi_\mu^\lambda(\alpha)\frac{\partial u_\lambda^i(x')}{\partial\alpha^\nu} = \frac{\partial\Psi_\mu^\lambda(\alpha)}{\partial\alpha^\nu}u_\lambda^i(x') - \frac{\partial\Psi_\nu^\kappa(\alpha)}{\partial\alpha^\mu}u_\kappa^i(x').$$

Replace the terms $\frac{\partial u}{\partial\alpha}$ appearing on the left by

$$\frac{\partial u_\kappa^i(x')}{\partial\alpha^\mu} = \frac{\partial x^j}{\partial x^j}\cdot\frac{\partial u_\kappa^\mu(x)}{\partial\alpha^\mu} = \frac{\partial x^j}{\partial\alpha^\mu}\cdot\frac{\partial u_\kappa^\mu(x)}{\partial x^j} = \Psi_\mu^\lambda(\alpha)u_\lambda^j(x)\cdot\frac{\partial u_\kappa^i(x)}{\partial x^j},$$

$$\frac{\partial u_\lambda^i(x')}{\partial\alpha^\nu} = \frac{\partial x^j}{\partial x^j}\cdot\frac{\partial u_\lambda^i(x)}{\partial\alpha^\nu} = \frac{\partial x^j}{\partial\alpha^\nu}\cdot\frac{\partial u_\lambda^i(x)}{\partial x^j} = \Psi_\nu^\delta(\alpha)u_\delta^j(x)\cdot\frac{\partial u_\lambda^i(x)}{\partial x^j}.$$

The rearranged equation becomes

$$\Psi_\nu^\kappa(\alpha)\Psi_\mu^\lambda(\alpha)u_\lambda^j(x)\cdot\frac{\partial u_\kappa^i(x)}{\partial x^j} - \Psi_\mu^\lambda(\alpha)\Psi_\nu^\delta(\alpha)u_\delta^j(x)\cdot\frac{\partial u_\lambda^i(x)}{\partial x^j} = \frac{\partial\Psi_\mu^\lambda(\alpha)}{\partial\alpha^\nu}u_\lambda^i(x') - \frac{\partial\Psi_\nu^\kappa(\alpha)}{\partial\alpha^\mu}u_\kappa^i(x').$$

Changing κ to δ on the LHS, and λ to κ on the RHS (which is okay since we didn't lose any terms with the new labels

$$\Psi_{\nu}^{\delta}(\alpha)\Psi_{\mu}^{\lambda}(\alpha)\left\{u_{\lambda}^{j}(x)\cdot\frac{\partial u_{\delta}^{i}(x)}{\partial x^{j}}-u_{\delta}^{j}(x)\cdot\frac{\partial u_{\lambda}^{i}(x)}{\partial x^{j}}\right\}=\left\{-\frac{\partial\Psi_{\nu}^{\kappa}(\alpha)}{\partial\alpha^{\mu}}+\frac{\partial\Psi_{\mu}^{\lambda}(\alpha)}{\partial\alpha^{\nu}}\right\}u_{\kappa}^{i}(x').$$

The terms $\Psi(\alpha)$ on the LHS can be moved to the RHS by multiplying both sides of the equation by the inverses of the $\Psi(\alpha)$, namely the $\Theta(\alpha)$. This renders the LHS only a function of x. The LHS "lives" in the differentiable manifold G. On the RHS is *almost* only a function of α. The RHS is almost an animal living in $\mathcal{T}$. The term $u_{\kappa}^{i}(x')$ is a function of x and cannot be moved to the LHS since it is an $\eta \times N$ matrix. We can circumnavigate this problem if we observe that these arguments hold for a Lie group of transformations. Isomorphic arguments hold for a Lie group. That is we can make the following identifications

$$f^{i}(\alpha,x)\to\phi'^{i}(\alpha,x),$$

$$u_{\lambda}^{i}(x)=\Theta_{\lambda}^{i}(\chi).$$

The shit happening in G is happening isomorphically in $\mathcal{T}$. (Go back to the example for Lie's first theorem and checkout the parallelism for yourself.) Then our almost separated example becomes

$$\Psi_{\nu}^{\delta}(\alpha)\Psi_{\mu}^{\lambda}(\alpha)\left\{\Theta_{\lambda}^{j}(\chi)\cdot\frac{\partial\Theta_{\delta}^{i}(\chi)}{\partial\chi^{j}}-\Theta_{\delta}^{j}(\chi)\cdot\frac{\partial\Theta_{\lambda}^{i}(\chi)}{\partial\chi^{j}}\right\}=\left\{-\frac{\partial\Psi_{\nu}^{\kappa}(\alpha)}{\partial\alpha^{\mu}}+\frac{\partial\Psi_{\mu}^{\lambda}(\alpha)}{\partial\alpha^{\nu}}\right\}\Theta_{\kappa}^{i}(\chi').$$

Let's multiply the equation by the inverses of $\Psi_{\nu}^{\delta}(\alpha)\Psi_{\mu}^{\lambda}(\alpha)$, and the inverse of $\Theta_{\kappa}^{i}(\chi')$.

$$\Psi_{i}^{\kappa}(\chi)\left\{\Theta_{\lambda}^{j}(\chi)\cdot\frac{\partial\Theta_{\delta}^{i}(\chi)}{\partial\chi^{j}}-\Theta_{\delta}^{j}(\chi)\cdot\frac{\partial\Theta_{\lambda}^{i}(\chi)}{\partial\chi^{j}}\right\}=\Theta_{\delta}^{\nu}(\alpha)\Theta_{\lambda}^{\mu}(\alpha)\left\{-\frac{\partial\Psi_{\nu}^{\kappa}(\alpha)}{\partial\alpha^{\mu}}+\frac{\partial\Psi_{\mu}^{\lambda}(\alpha)}{\partial\alpha^{\nu}}\right\}.$$

Since α and κ are arbitrary elements of $\mathcal{T}$ we have a "classic" separation of variables situation. Both sides must equal a constant, namely the equation $C_{\delta\lambda}^{\kappa}$. We have separation of variables:

$$\Theta_\lambda^j(\chi)\cdot\frac{\partial\Theta_\delta^i(\chi)}{\partial\chi^j}-\Theta_\delta^j(\chi)\cdot\frac{\partial\Theta_\lambda^i(\chi)}{\partial\chi^j}=C_{\delta\lambda}^\kappa\Theta_\kappa^i(\chi),$$

So back in the isomorphic differentiable manifold G, we now have nirvana.

$$u_\lambda^j(x)\cdot\frac{\partial u_\delta^i(x)}{\partial x^j}-u_\delta^j(x)\cdot\frac{\partial u_\lambda^i(x)}{\partial x^j}=C_{\delta\lambda}^\kappa u_\kappa^i(x').$$

1. For a Lie group of transformations (in the geometric manifold G) *with infinitesimal generators given by*

$$X_\lambda(x)=-\frac{\partial f^i(\beta,x)}{\partial\beta^\lambda}\bigg|_{\beta=0}\cdot\frac{\partial}{\partial x^i}=-u_\lambda^i(x)\frac{\partial}{\partial x^i}$$

the commutator relations are

$$[X_\lambda(x),X_\delta(x)]=\left[-u_\lambda^i(x)\frac{\partial}{\partial x^i},-u_\kappa^i(x)\frac{\partial}{\partial x^\kappa}\right]=C_{\delta\lambda}^\kappa\left(u_\kappa^j(x)\frac{\partial}{\partial x^j}\right)=C_{\delta\lambda}^\kappa X_\kappa(x).$$

2. For a Lie group with generators (in $\mathcal{T}$)

$$X_\lambda(\alpha)=-\frac{\partial\phi^\mu(\beta,\alpha)}{\partial\beta^\lambda}\bigg|_{\beta=0}\cdot\frac{\partial}{\partial\alpha^\mu}=-\Theta_\lambda^\mu(\alpha)\frac{\partial}{\partial\alpha^\mu}$$

the commutation relations are obtained in an isomorphic way:

$$[X_\lambda(\alpha),X_\delta(\alpha)]=\left[-\Theta_\lambda^\varepsilon(\alpha)\frac{\partial}{\partial\alpha^\varepsilon},-\Theta_\delta^\sigma(\alpha)\frac{\partial}{\partial\alpha^\sigma}\right]=\left[\Theta_\lambda^\varepsilon(\alpha)\frac{\partial\Theta_\delta^\sigma(\alpha)}{\partial\alpha^\varepsilon}-\Theta_\delta^\varepsilon(\alpha)\frac{\partial\Theta_\lambda^\sigma(\alpha)}{\partial\alpha^\varepsilon}\right]\frac{\partial}{\partial\alpha^\sigma}$$

$$=C_{\delta\lambda}^\kappa\left(-\Theta_\kappa^\sigma(\alpha)\frac{\partial}{\partial\alpha^\sigma}\right)=C_{\delta\lambda}^\kappa X_\kappa(\alpha).$$

We have now shown that the structure constants are constant for a Lie group, a Lie group of transformations, and for any of their analytic realizations or matrix representations. Q*uod Erat Demonstrandum dudes.*

Example 7—The PDE from example 6

$$\begin{pmatrix} \frac{\partial x'}{\partial \alpha^1} \\ \frac{\partial x'}{\partial \alpha^2} \end{pmatrix} = \begin{pmatrix} 1 & -\alpha^2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ 1 \end{pmatrix} = \begin{pmatrix} x' - \alpha^2 \\ 1 \end{pmatrix}$$

for x in terms of α are

$$\frac{\partial}{\partial \alpha^2} \cdot \frac{\partial x}{\partial \alpha^1} = \frac{\partial}{\partial \alpha^2}(x - \alpha^2) = \frac{\partial x}{\partial \alpha^2} - 1 = 1 - 1 = 0,$$

$$\frac{\partial}{\partial \alpha^1} \cdot \frac{\partial x}{\partial \alpha^2} = \frac{\partial}{\partial \alpha^1}(1) = 0.$$

Therefore, we do not violate the integrability condition that all mixed derivatives be equal

$$\frac{\partial^2 x'^i}{\partial \alpha^\mu \partial \alpha^\nu} = \frac{\partial^2 x'^i}{\partial \alpha^\nu \partial \alpha^\mu},$$

so we can find a unique solution to these PDEs with unique initial value $f(\alpha = 0; x_0) = x_0$.

The structure constants can be found from the group composition function ϕ. Since they are constant, they can be evaluated at any element. At the identity element $\Theta(\varepsilon) = \Psi(\varepsilon) = \mathbb{I}$, with $\chi = \mathbb{I}$ so that $\Theta^j_\lambda(\chi) = \Theta^j_\lambda(\mathcal{E}) = \mathbb{I}$, (but $\partial \chi^j \neq \varepsilon$ because it is a translation away from ε) our old equation

$$\Theta^j_\lambda(\chi) \cdot \frac{\partial \Theta^i_\delta(\chi)}{\partial \chi^j} - \Theta^j_\delta(\chi) \cdot \frac{\partial \Theta^i_\lambda(\chi)}{\partial \chi^j} = C^\kappa_{\delta\lambda} \Theta^i_\kappa(\chi)$$

Collapses to

$$C^\kappa_{\delta\lambda} = \frac{\partial \Theta^\kappa_\delta(\chi)}{\partial \alpha^\lambda} - \frac{\partial \Theta^\kappa_\lambda(\chi)}{\partial \alpha^\delta} = \frac{\partial^2 \phi^\kappa(\beta, \alpha)}{\partial \beta^\delta \partial \alpha^k} - \frac{\partial^2 \phi^\kappa(\beta, \alpha)}{\partial \beta^\lambda \partial \alpha^\delta}\bigg|_{\beta = \alpha = 0}.$$

For our continuing example with $\phi^1(\beta, \alpha) = \beta^1 + \alpha^1$, $\phi^2(\beta, \alpha) = e^{\beta^1} \alpha^2 + \beta^2$, we readily compute

$$C^2_{12} = \frac{\partial^2 \phi^2(\beta, \alpha)}{\partial \beta^1 \partial \alpha^2} - \frac{\partial^2 \phi^2(\beta, \alpha)}{\partial \beta^2 \partial \alpha^1} = 1.$$

Lie's third theorem. The structure constants obey

$$C^{\kappa}_{\delta\lambda} = -C^{\kappa}_{\lambda\delta},$$

$$C^{\sigma}_{\alpha\beta}C^{\rho}_{\sigma\gamma} + C^{\sigma}_{\beta\gamma}C^{\rho}_{\sigma\alpha} + C^{\sigma}_{\gamma\alpha}C^{\sigma}_{\sigma\beta} = 0.$$

Proof. From $C^{\kappa}_{\delta\lambda} = \frac{\partial^2\phi^{\kappa}(\beta,\alpha)}{\partial\beta^{\delta}\partial\alpha^{k}} - \frac{\partial^2\phi^{\kappa}(\beta,\alpha)}{\partial\beta^{\lambda}\partial\alpha^{\delta}}\Big|_{\beta=\alpha=0}$,

we have

$$C^{\kappa}_{\delta\lambda} = \frac{\partial^2\phi^{\kappa}(\beta,\alpha)}{\partial\beta^{\lambda}\partial\alpha^{\delta}} - \frac{\partial^2\phi^{\kappa}(\beta,\alpha)}{\partial\beta^{\delta}\partial\alpha^{k}}\Bigg|_{\beta=\alpha=0}.$$

The last equation is trivial, a trivial consequence of the Jacobi identity

$$\big[[X_\alpha, X_\beta]X_\gamma\big] + \big[[X_\beta, X_\gamma]X_\alpha\big] + \big[[X_\gamma, X_\alpha]X_\beta\big] = 0.$$

The Jacobi identity follows from the associativity of the group multiplication. The Jacobi identity bears a strong resemblance to

$$\frac{d}{dx}\big(f(x)g(x)\big) = \frac{df(x)}{dx}g(x) + \frac{f(x)dg(x)}{dx}.$$

This is why the Lie bracket [,] is sometimes called a derivative.

Example 8—Four of the eight structure constants for our elementary Lie group are necessarily zero by the antisymmetry requirement:

$$C^{1}_{ii} = C^{2}_{11} = C^{1}_{22} = C^{2}_{22} = 0.$$

The only remaining four structure constants are independent and must be computed explicitly:

$$C^{1}_{12} = -C^{1}_{21} = 0, \qquad C^{2}_{12} = -C^{2}_{21} = 1.$$

(Important to physicists). The structure constants for a Lie algebra provide a matrix representation for the algebra. Generally, this representation is unfaithful. The representation is obtained by associating with $X_\mu(\mu = 1,2, \dots, \eta)$ and $\eta \times \eta$ matrix $M_\mu(\eta \times \eta)$ whose matrix elements are given by $\left(M_\mu\right)_\alpha^\beta = -C_{\mu\alpha}^\beta$.

Theorem. $X_\mu \to M_\mu$, $\left(M_\mu\right)_\alpha^\beta = -C_{\mu\alpha}^\beta$ is a representation for the Lie algebra of the commutators X_μ.

Proof. Let's show that X and M have isomorphic commutator algebra.

$$\left(\left[M_\mu, M_\nu\right]\right)_\alpha^\beta = \left(M_\mu\right)_\alpha^\gamma (M_\nu)_\gamma^\beta - (M_\nu)_\alpha^\gamma \left(M_\mu\right)_\gamma^\beta = C_{\mu\alpha}^\gamma C_{\nu\gamma}^\beta - C_{\nu\alpha}^\gamma C_{\mu\gamma}^\beta.$$

The Jacobi identity tells us

$$C_{\alpha\mu}^\gamma C_{\gamma\nu}^\beta + C_{\mu\nu}^\gamma C_{\gamma\alpha}^\beta + C_{\nu\alpha}^\gamma C_{\gamma\mu}^\beta = 0.$$

$$C_{\mu\alpha}^\gamma C_{\nu\gamma}^\beta - C_{\nu\alpha}^\gamma C_{\mu\gamma}^\beta = C_{\mu\nu}^\gamma \left(-C_{\gamma\alpha}^\beta\right).$$

Then $\left[M_\mu, M_\nu\right]_\alpha^\beta = C_{\mu\nu}^\gamma \left(M_\gamma\right)_\alpha^\beta$. Since matrix multiplication is well defined, these matrices trivially satisfy the Jacobi identity; they also form a basis for a linear vector space. Since X_α and M_μ have is isomorphic commutation relations, the M_μ provide a representation of the X_μ.

Example 9—From our above structure constants, we construct a 2×2 matrix representation for our two-dimensional Lie group:

$$X_1 \to M_1 = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}, X_2 \to M_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Have we seen these before? **<u>Physicists</u>** this representation is sometimes called the ***adjoint*** or ***regular representation***, and it is used **extensively** in the study of the properties of Lie algebras, in particular when we seek to classify the various types of Lie algebras.

Converses of Lie's three theorems. Lie's three theorems show that for each Lie group there is a corresponding Lie algebra. These three theorems characterize the Lie algebra by its structure constants. The Lie algebra is uniquely determined, up to a change of basis, by the Lie group. Now we ask, given a Lie algebra, can we find an associated Lie group? No. There is generally not a 1-1 correspondence between Lie groups and algebras. Many Lie groups may have the same Lie algebra (recall Part I ODEs and PDEs with the same Lie group symmetries (as analytic realizations), but among all of the groups with the same Lie algebra there is only one that is simply connected. This simply connected group is the ***universal covering group***. I have yet to build up the little bit of algebraic topology required to deal with the converses of Lie's three theorems. The essence of the converses of Lie's three theorems is finding restricted cases for when one may uniquely associate a Lie algebra to a Lie group. The theorems are stated and proved in "An Introduction to the Lie Theory of One-Parameter Groups," Cohen, P. M., New York: Stechert, 1931. The converses of Lie's three theorems establish that there is a 1-1 correspondence between simply connected Lie groups and Lie algebras, but these converse theorems do not provide a method of constructing the analytic group multiplication $\gamma^{\mu} = \phi^{\mu}(\beta, \alpha)$ from the Lie algebra. Taylor's theorem goes to doing this construction in a canonical way. **Physicists**, Taylor's theorem lies at the heart of why we were able to find the time evolution of physical systems via iteration (exponentiation). It's extension underpins time ordering operators in path integral applications to quantum physics.

Example 10—Let's motivate the theorem with an example first. It helps. We keep with the chain of examples we have been using since example 1. For our two-dimensional group, this is a straight line: $\tau s^{\lambda} = \tau(\alpha^1, 0)$. (Recall: $\phi(\tau^1\tau^2, \alpha^1\alpha^2) = (\tau^1\alpha^1, \tau^1\alpha^2 + \tau^2)$, but we have $\alpha^2 = 0$, so $\tau(\alpha^1, 0) = (\tau^1\alpha^1, \tau^2)$. ($\tau s^{\lambda}$ is the name for the straight line; the RHS is the operational description of the line.) This is a straight line in the Lie group. Let

$$T(\tau = 1) \equiv e^{-\alpha^1 X_1(x)} = e^{-\alpha^1\left(-x\frac{\partial}{\partial x}\right)} = e^{\alpha^1 x\frac{\partial}{\partial x}} \approx 1 + \alpha^1 x\frac{\partial}{\partial x}.$$

Then it follows that

$$T(\tau = 1)x = \left(1 + \alpha^1 x \frac{\partial}{\partial x}\right) x = 1 + \alpha^1 x \sim e^{\alpha^1 x}.$$

Now let's repeat this for the straight line $\tau s^\lambda = \tau(0, \alpha^2)$.

$$T(\tau = 1) \equiv e^{-\alpha^2 X_2(x)} = e^{-\alpha^2\left(-\frac{\partial}{\partial x}\right)} \approx 1 + \alpha^2 \frac{\partial}{\partial x}.$$

Thus

$$T(\tau = 1)x = \left(1 + \alpha^2 \frac{\partial}{\partial x}\right) x = x + \alpha^2.$$

Lastly, consider the straight line $\tau s^\lambda = \tau(\alpha^1, \alpha^2)$. Then $T(\tau) = e^{\tau\left(\alpha^1 x \frac{\partial}{\partial x} + \alpha^2 \frac{\partial}{\partial x}\right)}$. So

$$T(\tau)x = e^{\tau\left(\alpha^1 x \frac{\partial}{\partial x} + \alpha^2 \frac{\partial}{\partial x}\right)} x = \left(1 + \tau\alpha^1 x \frac{\partial}{\partial x} + \tau\alpha^2 \frac{\partial}{\partial x}\right) x = (1 + \tau\alpha^1)x + \tau\alpha^2 = e^{\tau\alpha^1} x + \frac{\alpha^2}{\alpha^1}\left(e^{\tau\alpha^1} - 1\right).$$

I hate shenanigans. The source author didn't show the intermediate steps. $\frac{\alpha^2}{\alpha^1}\left(e^{\tau\alpha^1} - 1\right) \approx$ $\frac{\alpha^2}{\alpha^1}(1 + \tau\alpha^1 - 1)$. (This has a smell of bogusness.) But now $T(\tau)$ above compares well with $\tau(\alpha^1, \alpha^2) =$ $\begin{pmatrix} e^\wedge(\tau\alpha^1 & \frac{\alpha^2}{\alpha^1}\left(e^{\tau\alpha^1} - 1\right) \\ 0 & 1 \end{pmatrix}$. Now let's see about Taylor's theorem for Lie groups.

Taylor's theorem for Lie groups. There exists an analytic mapping $\gamma^\mu = \phi^\mu(\beta, \alpha)$ in which every straight line through the origin is a one-dimensional abelian subgroup. The Lie group operation corresponding to the Lie algebra element is $\alpha^\mu X_\mu \rightarrow e^{-\alpha^\mu X_\mu}$.

Proof. Since $X_\mu(x) = -u_\mu^j(x) \frac{\partial}{\partial x^j}$ (see Lie's 2nd theorem) we may write $\frac{\partial x^i}{\partial \alpha^\lambda} = \Psi_\lambda^\sigma u_\sigma^i(\alpha)$ as

$$\frac{\partial x^i}{\partial \alpha^\lambda} = -\Psi_\lambda^\sigma u_\sigma^i(\alpha) X_\sigma(x) x^i.$$

(Clever, since $-u^j_\mu(x)\frac{\partial}{\partial x^j}x^i$ leaves only $-u^i_\mu(x)$.) When we look at the straight line $\alpha^\mu(\tau) = s^\mu\tau$ through the origin of the Lie algebra, the x^i are functions of the single parameter τ:

$$\frac{dx^i(\tau)}{d\tau} = \frac{\partial x^i}{\partial\alpha^\lambda}\cdot\frac{d\alpha^\lambda}{d\tau} = -s^\lambda\Psi^\sigma_\lambda[\alpha = s\tau]X_\sigma(x)x^i(\tau).$$

Note that $s^\lambda = \frac{d\alpha^\lambda}{d\tau}$. Since the coordinates of the fixed point $[\left[x^i(p)\right]$ are now reduced to being functions of a single parameter τ, we can write

$$x^i(\tau) = x^j(0)T^i_j(\tau).$$

Then our differential equation $\frac{dx^i(\tau)}{d\tau}$ reduces to

$$\frac{d}{d\tau}T^i_j(\tau)x^j(0) = -s^\lambda\Psi^\sigma_\lambda[\alpha = s\tau]X_\sigma[x(\tau)]T^i_j(\tau)x^j(0).$$

Since the $x^i(0)$ are arbitrary, we have the following matrix equation

$$\frac{d}{d\tau}T^i_j(\tau) = -s^\lambda\Psi^\sigma_\lambda[\alpha = s\tau]X_\sigma[x(\tau)]T^i_j(\tau)$$

which is a first-order total differential equation for the matrix $T(\tau)$ with initial conditions $T^i_j(0) = \delta^i_j$.

Thus,

$$\frac{d}{d\tau}T^i_j(\tau)|_{\tau=0} = -s^\lambda\Psi^\sigma_\lambda[0]X_\sigma[x(0)]T^i_j(0) = -s^\lambda X_\sigma[x(0)]\delta^i_j.$$

(I went back to Lie's first theorem to see that $\Psi^\sigma_\lambda[0] = \mathbf{I}$ because its inverse $\Theta(0)$ is the identity.) The total differential equation has solution

$$T^i_j(\tau) = \sum_{n=0}^{\infty}\frac{1}{n!}\left[-\tau s^\lambda X_\lambda[x(0)]\delta^i_j\right]^n = e^{-\tau s^\lambda X_\lambda(x)\delta^i_j} = \delta^i_j e^{-\alpha^\lambda X_\lambda(x)}.$$

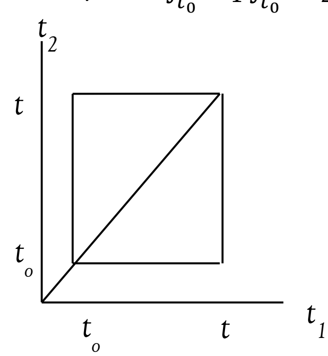
The solution is unique and analytic if it converges. Also, since

$$e^{-\tau_1 s^\lambda X_\lambda} e^{-\tau_2 s^\lambda X_\lambda} = e^{-(\tau_1+\tau_2) s^\lambda X_\lambda},$$

every straight line through the origin of the algebra maps into a one-dimensional abelian subgroup. Now we can go back to example 10 to see this theorem's application with a fuller appreciation.

Theorem (without proof) Every element of a compact Lie group G lies on a one-dimensional abelian subgroup of G and can be obtained by exponentiating some element of the Lie algebra. (Often theorems that hold for compact groups fail for noncompact groups ($SU(N)$ are compact; the Poincaré group is noncompact.)

On time ordered operators (and Taylor's theorem for Lie groups). Using the figure below $\int_{t_0}^{t} dt_1 \int_{t_0}^{t} dt_2$ covers the entire area as of the square. $\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2$ only covers lower half triangular area. (Think of the *t*'s as time.)



So,

$$\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2\, H(t_1)H(t_2) = \frac{1}{2}\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2\, H(t_1)H(t_2).$$

Also, flipping the order on the LHS,

$$\int_{t_0}^{t} dt_2 \int_{t_0}^{t_1} dt_1\, H(t_1)H(t_2) = \frac{1}{2}\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} dt_2\, H(t_1)H(t_2).$$

The mathematics is oblivious to the ordering of the integrals. To motivate τ-ordered products, let me quote from the quantum mechanical formulation of path integrals. Path integrals in physics are set up with time proceeding from earlier times to later times. If $t_2 > t_1$

$$\langle q_f t_f | q(t_1) q(t_2) | q_i t_i \rangle = \int \frac{\mathcal{D}q\mathcal{D}p}{\hbar} q(t_1) q(t_2) e^{\frac{i}{\hbar}\int_{t_1}^{t_2}(p\dot{q}-\mathcal{H})dt},$$

but if $t_1 > t_2$, then $\langle q_f t_f | q(t_2) q(t_1) | q_i t_i \rangle$. This violation of time flow is handled in physics by the time ordering operator

$$T[A(t_1)B(t_2)] = \begin{Bmatrix} A(t_1)B(t_2) & if\ t_1 > t_2 \\ B(t_2)A(t_1) & if\ t_2 > t_1 \end{Bmatrix}.$$

Let's get back to Taylor's theorem for Lie groups. The theorem involved integration in the Lie algebra along a straight line from the origin to some point $\alpha = (\alpha^1, \alpha^2, \dots, \alpha^\eta)$. It is not necessary to integrate along a straight line. A curve will also work. However, to make the theorem work, we will see that τ-ordering of the group elements will be necessary.

Let $\alpha^\mu(\tau)$ be a curve in the algebra with the properties

$$\alpha^\mu(\tau = 0) = (0,0,\dots,0),$$

$$\alpha^\mu(\tau = 1) = (\alpha^1, \alpha^2, \dots, \alpha^\eta).$$

Then the differential equation

$$\frac{d}{d\tau} T_j^i(\tau) x^j(0) = -s^\lambda \Psi_\lambda^\sigma[\alpha = s\tau] X_\sigma[x(\tau)] T_j^i(\tau) x^j(0)$$

may be solved by iteration. (Along the straight line we had, notation-wise, $\alpha^\mu(\tau) = \tau\alpha^\mu$. On the curve, notation-wise, we have some α^μ, the "straight line" slope of which is $-\frac{d\alpha^\mu}{d\tau}$.) So, solving by iteration:

$$\frac{d}{d\tau}T(\tau) = -\frac{d\alpha^{\lambda}(\tau)}{d\tau}\Psi_{\lambda}^{\sigma}[\alpha(\tau)X_{\sigma}(x)T(\tau).$$

$$T(\tau) - T(0) = \int_0^{\tau} d\tau' \left\{-\frac{d\alpha^{\lambda'}(\tau')}{d\tau'}\Psi_{\lambda'}^{\sigma'}[\alpha(\tau')X_{\sigma'}(x)\left\{T(0)\right.\right.$$

$$+\int_0^{\tau'} -\frac{d\alpha^{\lambda''}(\tau'')}{d\tau''}\Psi_{\lambda''}^{\sigma''}[\alpha(\tau'')X_{\sigma''}(x)\left\{T(0)\right.$$

$$+\int_0^{\tau''} -\frac{d\alpha^{\lambda'''}(\tau''')}{d\tau'''}\Psi_{\lambda'''}^{\sigma'''}[\alpha(\tau''')X_{\sigma'''}(x)\{\ldots$$

Things simplify if we observe that $T(0) = \mathbf{I}$. Also, relabel the integrand

$$-\frac{d\alpha^{\lambda}(\tau)}{d\tau}\Psi_{\lambda}^{\sigma}[\alpha(\tau)X_{\sigma}(x) = -\frac{i}{\hbar}\mathcal{H}(\tau).$$

(The mathematicians are sneaking in the physicist's Hamiltonian.) (Here is where τ-ordering (or time ordering) comes in.) Note that the term coming from the second iteration can be rewritten

$$\int_0^{\tau} d\tau' \left(-\frac{i}{\hbar}\right)\mathcal{H}(\tau')\int_0^{\tau} d\tau'' \left(-\frac{i}{\hbar}\right)\mathcal{H}(\tau'')$$

$$= \frac{1}{2!}\left\{\left(-\frac{i}{\hbar}\right)^2 \int_0^{\tau} d\tau'\,\mathcal{H}(\tau')\int_0^{\tau} d\tau''\,\mathcal{H}(\tau'') + \left(-\frac{i}{\hbar}\right)^2 \int_0^{\tau} d\tau' \int_{\tau'}^{\tau} d\tau''\,\mathcal{H}(\tau'')\mathcal{H}(\tau')\right\}$$

$$= \frac{1}{2!}\tau\left(-\frac{i}{\hbar}\right)^2 \int_0^{\tau}\int_0^{\tau} d\tau' d\tau'' \mathcal{H}(\tau'')\mathcal{H}(\tau'),$$

where the symbol τ symbolizes an ordering operator:

$$\tau\mathcal{H}(\tau')\mathcal{H}(\tau'') = \begin{cases}\mathcal{H}(\tau')\mathcal{H}(\tau'') \;\; if\; \tau' > \tau'' \\ \mathcal{H}(\tau'')\mathcal{H}(\tau') \;\; if\; \tau'' > \tau'\end{cases}.$$

(The 2! Is counting upper/lower triangular areas.) If we define

$$\theta(x) = \int_{-\infty}^{x} \delta(y)dy = \begin{cases} 1, & x > 0 \\ \frac{1}{2}, & x = 0, \\ 0, & x < 0 \end{cases}$$

then we can write

$$\tau\mathcal{H}(\tau')\mathcal{H}(\tau'') = \mathcal{H}(\tau')\mathcal{H}(\tau'') + \theta(\tau'' - \tau')[\mathcal{H}(\tau''), \mathcal{H}(\tau')].$$

Clearly the τ-ordered product is not equal to the usual product unless the commutator vanishes. Third and higher order iterations can be treated analogously. The iterative solution is then

$$T(\tau) = \tau e^{\left(-\frac{i}{\hbar}\int_0^{\tau}\mathcal{H}(\tau')d\tau'\right)}.$$

This result collapses to the straight line solution over a straight line (see Gilmore's Dover text.) I personally understand every section in the notes well except for this section. The physicist's version of all of this for path integrals is far more sensible to me—apologies.

Topological concepts. Recall that a space is connected if any two points in the space can be joined by a line (in the sense of "line" we used to prove Taylor's theorem for Lie groups), and all points of the line lie in the space.

Example 11—Still keeping with our running example, the topological space $\mathcal{T}_2$ (I've been lazily writing $\mathcal{T}$) is not connected because any line joining, say (1,0) and (-1,0) must contain a point $(\alpha^1 = 0, \alpha^2)$ which is not in $\mathcal{T}_2$. If you go back to example 1, you'll recall we had $x'(p) = -x(p)$. Thus $\mathcal{T}_2 = \mathbb{R}^2 - (\alpha^1 = 0, \alpha^2)$. The y axis splits $\mathbb{R}^2$.

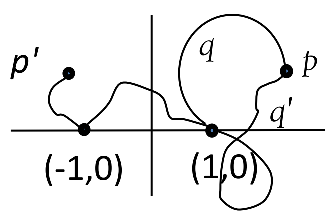
If p is any point, we can look at the set of points connected to p, *e.g.*, all points in the left half-plane are connected to (-1,0); all points in the right half-plane are connected to (1,0). A connected component of a continuous group (called a **sheet**) cannot itself be a group unless it contains the identity—duh.

Theorem—The component of a continuous group that is connected with the identity is a group.

In addition, all other sheets of the continuous group are isomorphic as manifolds, both to each other and to the connected component. A new group structure can also be defined on each sheet, so that it becomes isomorphic with the connected component.

A space is ***simply connected*** if a curve connecting any two points in the space can be continuously deformed into every other curve connecting the same two points.

Example 12—The connected component of $G_{(1,0)}$ is simply connected since any curve $(1,0) - q - p$ can be continuously deformed to any other curve $(1,0) - q' - p'$. (Recall the vertical axis is not in $\mathcal{T}_2$.



You cannot get to the identity element from the left hand plane. The left hand plane is not simply connected.

Algebraic concepts. This section assumes a background in abstract algebra up to the concepts of cosets and normal subgroups. Don't fret if you don't have this background. I build it up, step-by-step in the first few pages of Part V. Try reading the section first. I think you might already have sufficient background and context from our discussion of commutators to get this section.

Let H be a subgroup of G (in our mind G being a continuous Lie group). Then for any element $g \epsilon G$, the set of group operations $gHg^{-1} = \{gh_i g^{-1}, h_i \epsilon H\}$ also form a group, the ***group conjugate to*** H. (Think of gHg^{-1} in the sense of a commutator measuring how much the algebra may differ from being commutative; I never got this in junior, senior and graduate courses in abstract algebra.)

Example 13—The one-dimensional subgroup of $G_{(1,0)}$ (the right hand half-plane of example 1) given by $(\alpha^1 > 0, \alpha^2 = 0)$ gives rise to the series of conjugate groups

$$(\beta^1,\beta^2)(\alpha^1,0)(\beta^1,\beta^2)^{-1} = (\beta^1\alpha^1,\beta^1\cdot 0+\beta^2)(\beta^1,\beta^2)^{-1} = (\beta^1\alpha^1,\beta^2)\left(\frac{1}{\beta^1},-\frac{\beta^2}{\beta^1}\right)$$

$$= (\alpha^1,-\alpha^1\beta^2+\beta^2) = \left(\alpha^1,\beta^2(1-\alpha^1)\right).$$

You can check the group axioms to confirm that the subset of group elements $\left(\alpha^1,\beta^2(1-\alpha^1)\right)$ generated by conjugation is indeed a subgroup of $\mathcal{T}_2$ corresponding to $G_{(1,0)}$. The subgroup conjugate to the abelian group $(1,\alpha^2)$ is

$$(\beta^1,\beta^2)(1,\alpha^2\ )(\beta^1,\beta^2)^{-1} = (1,\beta^1\alpha^2).$$

Every element in this conjugate subgroup is in the original subgroup. We signify this by writing $gHg^{-1} = H$. Subgroups H that are self-conjugate for all elements of $g\epsilon G$, are called ***invariant subgroups*** or ***normal subgroups***. The subgroup $H = (1,\alpha^2)$ is an abelian invariant subgroup $H\underline{\Delta}G$.

Theorem—The connected component G_0 of a continuous group is an invariant subgroup of G; $G_0\underline{\Delta}G$. (Welcome to some algebraic topology.)

Example 14—The right hand half-plane $G_{(1,0)}\underline{\Delta}G$.

If H is a subgroup of G, $(H \leq G)$, the structure G/H are called ***cosets***. A right coset is the set of group operations $c_0, c_1, \dots, \epsilon G\ \ni$ (∋ reads such that)

$$Hc_0 + Hc_1 + \cdots = G,$$

where no element $g\epsilon G$ is contained in more than once in the sum above as $Hc_i \cap Hc_j = \emptyset$ if $i \neq j$. The right (and left) cosets C_R are disjoint. In another way of saying it, every element $g\epsilon G$ can be written uniquely (!)

$$g = h_i c_j, \qquad h_i \epsilon H, \qquad c_j \epsilon C_R.$$

Analogously, left cosets C_L involve a ! decomposition

$$g = c_k h_l, \qquad h_l \epsilon H, \qquad c_k \epsilon C_L.$$

Notation: A subgroup H of G is denoted by $H \leq G$. If $H \neq G$, it is a proper subgroup denoted by $H < G$.

Example 15—Let $G = G_{(1,0)}$. $H_1 = (\alpha^1 > 0, \alpha^2 = 0) < G$. $H_2 = (1, \alpha^2) < G$. (Geometrically, H_1 is the line $(0, \infty)$ lying on the α^1 axis; H_2 is a vertical line α^2 from $-\infty$ to ∞ situated at $\alpha^1 = 1$.)

$$(\gamma^1, \gamma^2) \overset{!}{=} (\gamma^1, 0) \circ \left(1, \frac{\gamma^2}{\gamma^1}\right).$$

We can read this equation in two ways: $(\gamma^1, \gamma^2) \in \mathrm{G} = \mathrm{H}_1 \circ C_R$ since $(\gamma^1, 0) \epsilon H$ and $\left(1, \frac{\gamma^2}{\gamma^1}\right) \epsilon\, C_R$, or $(\gamma^1, \gamma^2) \in \mathrm{G} = \mathrm{C_L} \circ H_2$ since $(\gamma^1, 0)\ \epsilon\ C_L$ and $\left(1, \frac{\gamma^2}{\gamma^1}\right)\ \epsilon\ H_2$. Draw pictures to better see this! It also never hurts to check things. Let's check $g h_2 g^{-1}$.

$$(\gamma^1, \gamma^2)(1, \alpha^2)(\gamma^1, \gamma^2)^{-1} = (\gamma^1, \gamma^2)(1, \alpha^2)\left(\frac{1}{\gamma^1}, -\frac{\gamma^2}{\gamma^1}\right) = (\gamma^1, \gamma^2)\left(\frac{1}{\gamma^1}, -\frac{\gamma^2}{\gamma^1} + \alpha^2\right)$$

$$= (1, -\gamma^2 + \gamma^1 \alpha^2 + \gamma^2) = (1, \gamma^1 \alpha^2)\ \epsilon\ H_2.$$

Theorem—Left and right cosets G/H of a continuous group G by a closed subgroup of H are manifolds of dimension $\dim G - \dim H$. (Wow).

I offer no proof here, but I do provide an example. We have

$$\dim G = \dim G_{(1,0)} = 2 - \dim H_1 = 1,$$

$$\dim G = \dim G_{(1,0)} = 2 - \dim H_2 = 1.$$

Theorem—If H is an invariant (normal) subgroup of G then the coset elements $c_0, c_1, \ldots$, can be chosen in such a way that they are closed under multiplication and form a group called the ***factor group*** G/H.

Note that H_1 is not an invariant subgroup of G, yet G/H_1 is a group. Accidents like this can happen. On the other hand $H_2 \underline{\Delta} G$ and the theorem above guarantees that $G.H_2$ is a group, a factor group.

Example 16—The connected component of G ($G_{(1,0)}$) is an invariant (normal) subgroup. Every element of G can be written, for example, as either

$$g \circ \left(\frac{1}{2}, 0\right) \quad \text{or} \quad g \circ (-\frac{1}{2}, 0)$$

$$g \in G_{(1,0)}, \qquad c_0 = \left(\frac{1}{2}, \right), \qquad c_1 = \left(-\frac{1}{2}, 0\right).$$

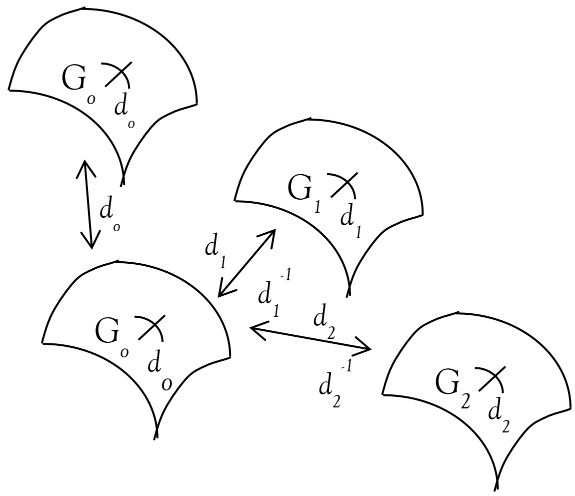
The coset representatives c_0, c_1 consist of one element from each sheet. This is not the most clever choice (an art) for coset representations. A better choice would be $c_0 = (1,0)$ and $C_1 = (-1,0)$. This forms a closed group ($\langle \mathbb{Z}_2, + \rangle$) since $(1,0)^2 = (1,0)$, $(-1,0)^2 = (1,0)$, $(1,0)(-1,0) = (-1,0) = (-1,0)(1,0)$. Generally, a good choice for coset representatives for the factor group

$$\frac{G_{\text{continous}}}{G_{\text{connected component}}}$$

would be the element in each sheet that becomes the identity under the mapping that turns that sheet into a group isomorphic (≅) with the connected component. Let's illustrate this. A many sheeted continuous group G can be written as the sum of group operations of the form

$$d_i G_0 = \bigcup_{i=0} d_i G_0,$$

where the $d_i \in D$ and G_0 is the connected component. Each sheet is topologically equivalent as a manifold. Only $G_0 = d_0 G_0$ can be made isomorphic with the subgroup G_0 by acting with d_i^{-1} on the left.



Theorem—The factor group of a continuous group by its connected component G_0 is a discrete group $D$of dimension 0. That is $G/G_0 = D$ and $\dim D = 0$.

With this theorem we can study all continuous groups if we study separately (A) only the connected continuous groups and (B) discrete groups.

Referring to the figure above, if we know the structure of the connected component G_0 of a topological (continuous) group, the structure of the discrete group $D = G/G_0$, and the structure of the mapping $DG_0D^{-1} \rightarrow G_0$, the structure of the entire continuous groups G can be constructed as follows. Let

$$g = \epsilon\, G \underset{\rightarrow}{!} d_i\alpha \qquad d_i, d_j\ \epsilon\ D\ \subset G.$$

$$h = \epsilon\, G \underset{\rightarrow}{!} d_i\beta \qquad \alpha, \beta\ \epsilon\ G_0\ \subset G.$$

Then

$$g \circ h = d_i\alpha \circ \alpha d_j\beta = d_i d_j\left(d_j^{-1}\alpha d_j\right)\beta = \left(d_k = d_i d_j\right)\left(\alpha' = d_j^{-1}\alpha d_j\right)\beta.$$

Since D has a group structure, the product $d_i \circ d_j = d_k \epsilon D$. Since G_0 is an invariant (normal) subgroup , $d_j^{-1} \alpha d_j \epsilon\ G_0$. Therefore, the group multiplication properties in the discrete factor group D and in the connected continuous component G_0, together with the mapping $\alpha \to \alpha' = d_j^{-1} \alpha d_j$ uniquely determine the group multiplication properties of the entire group G.

Let the elements of the discrete group D be $d_0 = \text{identity}, d_1, d_2, \dots, d_n, \dots$ and let G_i be the sheet containing d_i. Then the sheet containing the identity $d_0 \epsilon D$ is the connected component G_0. The mapping that converts any sheet G_i (which cannot be a group unless $i = 0$) into a group isomorphic with G_0 is given by

$$G_i = d_i G_0 \overset{d_i^{-1}}{\longrightarrow} d_i^{-1} d_i G_0 = G_0.$$

See the figure above for this process.

The invariant integral. The study of functions defined on the group manifold G require the use of an integral function defined over the group elements. <u>The rearrangement property</u>. The operation of multiplication by a group element may be interpreted either as a mapping of the entire topological group onto itself, or as a change of basis within the topological space. If $f(\beta)$ is any scalar-valued function defined on the group (or equivalently on the topological space $\mathcal{T}$) then a reasonable requirement for a group integral is

$$\int f(\beta) d\mu(\beta) = \int f(\alpha\beta) d\mu(\beta) = \int f(\alpha\beta) d\mu(\alpha\beta).$$

This is reasonable because if the sum over β involves each β exactly once, the sum over $\alpha\beta$ involves each group operation also exactly once (in a scrambled order). Since f is an arbitrary function, we demand of the measure defined on the group that it obey

$$d\mu_L(\beta) = d\mu_L(\alpha \circ \beta).$$

Measures with this property are called left invariant measures. Right invariant measures can also be defined: $d\mu_R(\beta) = d\mu_R(\beta \circ \alpha)$.

Reparameterization of G_0. We first define group integration on the connected component of a continuous group; then we extend it to the entire continuous group. We shall also demonstrate these concepts on the connected component of our old example group.

Since it's common to associate the origin of R_η with the group identity ε, reparameterization of the connected component G_0 may proceed as follows:

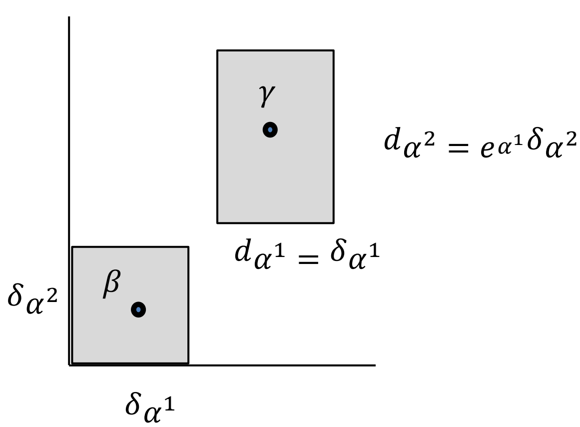
$$(\alpha^1, \alpha^2) \leftrightarrow \begin{pmatrix} e^{\alpha^1} & \alpha^2 \\ 0 & 1 \end{pmatrix}.$$

Under this reparameterization,

$$(\beta^1, \beta^2) \circ (\alpha^1, \alpha^2) = \left(\beta^1 + \alpha^2, e^{\beta^1} \alpha^2 + \beta^2\right),$$

$$(\alpha^1, \alpha^2)^{-1} = \left(-\alpha^1, -e^{-\alpha^1}\, \alpha^2\right).$$

Now let $\delta\alpha^1$, $\delta\alpha^2$ denote infinitesimal displacements in the topological space R_2. The volume at the identity enclosed by $\delta\alpha^1 \wedge \delta\alpha^2$ is shaded in the figure below (left invariant measure).



Let's fill in the details for the left invariant measure.

$$(\alpha^1, \alpha^2) \circ (0,0) = \left(\alpha^1 + 0, e^{\alpha^1} \cdot 0 + \alpha^2\right) = (\alpha^1, \alpha^2).$$

$$(\alpha^1, \alpha^2) \circ (\delta\alpha^1, 0) = \left(\alpha^1 + \delta\alpha^1, e^{\delta\alpha^1} \cdot 0 + \alpha^2\right) = (\alpha^1 + \delta\alpha^1, \alpha^2)$$

Since $\alpha^1 \to \alpha^1 + \delta\alpha^1$, then $d\alpha^1 = \delta\alpha^1$. There is no change of scale alone the α^1 axis. On the other hand,

$$(\alpha^1, \alpha^2) \circ (0, \delta\alpha^2) = \left(\alpha^1, e^{\alpha^1}\delta\alpha^2 + \alpha^2\right),$$

So $d\alpha^2 = e^{\alpha^1}\delta\alpha^2$. The transformed rectangle is stretched along the α^2 axis. Then the volume element $\delta V_L(\varepsilon) = d\alpha^1 d\alpha^2 = e^{\alpha^1}\delta\alpha^1\delta\alpha^2$.

What about the right invariant measure?

$$(0,0) \circ (\alpha^1, \alpha^2) = (0 + \alpha^1, e^0\alpha^2 + 0) = (\alpha^1, \alpha^2).$$

$$(\delta\alpha^1, 0) \circ (\alpha^1, \alpha^2) = \left(\alpha^1 + \delta\alpha^1, e^{\delta\alpha^1}\alpha^2 + 0\right) = \left(\alpha^1 + \delta\alpha^1, e^{\delta\alpha^1}\alpha^2\right).$$

So again $\alpha^1 \to \alpha^1 + \delta\alpha^1$, and $d\alpha^1 = \delta\alpha^1$. Now,

$$(0, \delta\alpha^2) \circ (\alpha^1, \alpha^2) = (0 + \alpha^1, e^0\alpha^2 + \delta\alpha^2) = (\alpha^1, \alpha^2 + \delta\alpha^2).$$

So $\alpha^2 \to \alpha^2 + \delta\alpha^2$. This seems right, but it is not what the source author concluded. (Shenanigans are afoot.) Here is what probably makes the most sense for extracting these deltas *the right way*. **Do it all in one shot**. (Doing it in pieces might get you lucky—the right results.)

$$(\delta\alpha^1, \delta\alpha^2) \circ (\alpha^1, \alpha^2) = \left(\alpha^1 + \delta\alpha^1, e^{\delta\alpha^1}\alpha^2 + \delta\alpha^2\right).$$

We get $\alpha^1 \to \alpha^1 + \delta\alpha^1$. Expand the exponential to first order, and $\alpha^2 \to \alpha^2 + \alpha^2\delta\alpha^1 + \delta\alpha^2$. So $d\alpha^2 = \alpha^2\delta\alpha^1 + \delta\alpha^2$. Fixing crap like this over spans of years and books is wasted time.

Let β be any point infinitesimally near to the identity and within the volume element δVas we have shown in the figure for left invariant measure (and our imaginary picture for right invariant measure). Under translation by $\alpha \circ (\alpha \circ \beta)$ the volume element $\delta V(\varepsilon)$ around ε expands as it is moved to a volume element $\delta V_L(\varepsilon)$ around α. The coordinates of any point $\alpha \circ (\delta\alpha^1 \circ \delta\alpha^2)$ may be expanded as

$$(\alpha + d\alpha)^\mu = \phi(\alpha, \delta\alpha) = \phi^\mu(\alpha, 0) + \left.\frac{\partial \phi^\mu(\alpha, \beta)}{\partial \beta^\lambda}\right|_0 \delta\alpha^\lambda + \cdots$$

The Euclidean volume element $\delta V(\varepsilon)$ expands as it is moved around by the factor

$$\alpha \circ \delta V(\varepsilon) = \delta V_L(\alpha) = d\alpha^1 \wedge d\alpha^2 = (\delta\alpha^1) \wedge \left(e^{\alpha^1}\delta\alpha^2\right)\delta V(\varepsilon).$$

(For the reader who has notice the operator ∧, I'm pretty sure the source author is thinking in terms of differential forms. We're okay thinking of it as a simple product in this work). Therefore we should weigh the Euclidean volume element in the vicinity of α by a factor $e^{-\alpha^1}$to preserve volume invariance over the entire topological space:

$$\text{Invariant volume element} = \text{density} \times \text{Euclidean volument elelment.}$$

For the right translation by α, we have, analogously,

$$\delta V(\varepsilon) \circ \alpha = \delta V_R(\alpha) = d\alpha^1 \wedge d\alpha^2 = (\delta\alpha^1) \wedge (\alpha^2\delta\alpha^1 + \delta\alpha^2) = \delta\alpha^1 \wedge \delta\alpha^2 = \delta V(\varepsilon).$$

We dropped $\delta\alpha^1 \wedge \alpha^2\delta\alpha^1$ because this is a 2nd order term. Evidently then, the right invariant integral is defined by a uniform density, whereas the left invariant density is nonuniform, defined by $e^{-\alpha^1}$.

General right and left invariant densities. We now turn to the general case. Let $\delta\alpha^1, \delta\alpha^2, \dots, \delta^{\alpha^\eta}$ be infinitesimal displacement n the η independent directions of R_η at the identity element 0. The volume they enclose is give by

$$\delta V(0) = \rho(0)\delta\alpha^1 \wedge \delta\alpha^2 \dots \delta\alpha^\eta.$$

Move this volume element to an arbitrary element $dV(0)$ to an arbitrary group element α by left translation, and demand of the density function that

$$\rho(0)dV(0) = \rho_L(\alpha)dV_L.$$

Here $dV_L(\alpha)$ is the Euclidean volume element $dV(0)$ after it has been moved from the vicinity of (0) to the vicinity of (α) by left translation with α. It is easy to compute dV_L:

$$dV_L(\alpha) = d\alpha^1 \wedge d\alpha^2 \wedge \dots \wedge d\alpha^\mu,$$

$$(\alpha + d\alpha)^\mu = \phi^\mu(\alpha,\backslash\delta\alpha) = \phi^\mu(\alpha, 0) + \left.\frac{\partial\phi^\mu(\alpha,\beta)}{\partial\beta^\lambda}\right|_{\beta=0.}$$

Then we have

$$dV_L(\alpha) = \left.\frac{\partial\phi^1(\alpha,\beta)}{\partial\beta^{\lambda^1}}\right|_{\beta=0} \delta\alpha^{\lambda^1} \wedge \left.\frac{\partial\phi^2(\alpha,\beta)}{\partial\beta^{\lambda^2}}\right|_{\beta=0} \delta\alpha^{\lambda^2} \wedge \dots \wedge = \det\left\|\frac{\partial\phi^\mu(\alpha,\beta)}{\partial\beta^\lambda}\right\|_{\beta=0} dV(0).$$

The density for left translations must thus satisfy $(\rho(0) = 1)$

$$\rho_L(\alpha) = \left\|\frac{\partial\phi^\mu(\alpha,\beta)}{\partial\beta^\lambda}\right\|_{\beta=0}^{-1}.$$

Doing all of the above for the right leads to

$$\rho_R(\alpha) = \left\|\frac{\partial\phi^\mu(\beta,\alpha)}{\partial\beta^\lambda}\right\|_{\beta=0}^{-1}.$$

These invariant densities are ***Haar measures***.

Equality of left and right measures. The left and right measures computed in our previous figure were not equal. Furthermore, the integrals

$$\int \rho_L(\alpha)dV(\alpha), \qquad \int \rho_R(\alpha)V(\alpha)$$

for our example do not converge (though any integral $\int f(\alpha)\rho_{L,R}(\alpha)\, dV(\alpha)$

which did converge would be invariant. Where are the left and right measures equal? And we do they converge?

Theorem—The density functions $\rho_L(\alpha)$ and $\rho_R(\alpha)$ giving the left and right invariant measures of compact group are equal.

Proof. Begin with an element of volume $dV(0)$ at the identity of the group operation. Move this element to an invariant volume at α by left translation. The result is

$$d\mu_L(\alpha) = \rho_L(\alpha)dV_L(\alpha).$$

If we now move this invariant volume back to the identity using right translation by α^{-1}, the result is a volume element at the identity, but with a possibly different size and shape. The left translation mapped a rectangle to a rectangle with the same orientation, but the right translation, had I drawn it, results in mapping a rectangle to a rotated parallelogram. Mathematically we have:

$$[\rho_L(\alpha)dV_L(\alpha) \circ \alpha^{-1} = \rho_L(\alpha)d\mu(0)\rho_R(\alpha^{-1})$$

$$d\mu^1(0) = [\rho_L(\alpha)\rho_R^{-1}(\alpha)]d\mu(0)$$

$$d\mu^1(0) = f(\alpha)d\mu(0).$$

Then define recursively

$$d\mu^{\eta+1}(0) = \alpha \circ d\mu^{\eta}(0) \circ \alpha^{-1} = (\alpha)^{\eta+1} \circ d\mu(0) \circ (\alpha^{-1})^{\eta+1}$$

$$d\mu^{\eta+1}(0) = f(\alpha) d\mu^{\eta}(0) = f^{\eta+1}(\alpha) d\mu(0).$$

Since the group is compact (and every sequence has a convergent subsequence) the limit as $n \to \infty$ exists and is well defined: $\lim_{n\to\infty}(\alpha)^n = \beta$, and

$$d\mu^{\infty}(0) = \lim_{n\to\infty} \alpha^{\mu} d\mu(0)(-\alpha)^n = \lim_{n\to\infty} [f(\alpha)]^n d\mu(0) = f(\beta) d\mu(0).$$

If $f(\alpha) \gtrless 1$, then $\lim_{n\to\infty} [f(\alpha)]^n$ is infinity for the > case and 0 for the < case. But since β is an element of the group, $f(\beta)$ can be neither 0 nor infinity, for then the operation $\beta \circ d\mu(0) \circ \beta^{-1}$ would be singular. We are forced to conclude that $f(\beta) = 1$, thus $\rho_L(\alpha) = \rho_R(\alpha)$ and $d\mu_L(\alpha) = d\mu_R(\alpha)$. The left and right measures on a compact group are equal.

Note from source author: The left and right measures equal on the following kinds of groups:

1. Finite groups.
2. Discrete groups.
3. (Locally compact) abelian groups.
4. Compact groups.
5. Simple groups.
6. Semisimple groups.
7. Real connected algebraic groups with determinant +1.
8. Connected nilpotent Lie groups.
9. Semidirect product groups such as $SO(3) \wedge T_3$ or $ISO(3)$; Poincaré groups $SO(3) \wedge T_{3,1}$, or ISO(3,1).
10. Constructions of semisimple groups by maximal subgroups: $ISO(p,q)$, $IU(p,q)$, $IUSp(2p,2q)$.
11. Lie groups for which $Ad(G)$ is unimodular.

Extension to continuous groups. The invariant integrals have so far been built for the connected components of a continuous group. To extend to the other sheets we need to observe two things. 1. Every element $\alpha \epsilon G$ can be written as $\alpha = d_i \circ \alpha_0$ where $d_i \epsilon D$, a discrete factor group, and $\alpha_0 \epsilon G_0$, the connected component. 2. Each sheet is topologically equivalent to every other sheet. The required extension is then

$$\int_{\alpha \, \epsilon \, G} f(\alpha)\rho(\alpha)dV(\alpha) = \sum_{d_i \, \epsilon \, D} \int_{\alpha_0 \, \epsilon \, G_0} f(d_i \circ \alpha_0)\rho(\alpha_0)dV(\alpha_0).$$

If it happens that the connected component is of dimension zero, it contains only one point, the identity and the invariant integral reduces to (with $\rho(\varepsilon) = 1$) to the form familiar for discrete groups:

$$\int f(\alpha)\rho(\alpha)dV(\alpha) \to \sum_{d_i \, \epsilon \, G} f(d_i).$$

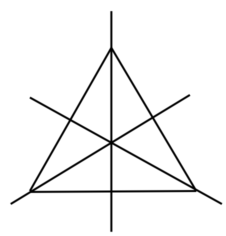
Done on this.

Part III were my notes from Gilmore's text, specifically from chapters 3 and 4, interspersed with filled in steps and underlying physical motivations. The first two chapters are worth a read, especially the stuff dealing with metrics. Three more blocks of ideas remain regarding symmetry methods. Chapters 1-4 of Jones' text develop the mathematics of discrete groups (our focus has been primarily on continuous groups, but not exclusively). We learn how to create characteristic tables for discrete groups. You connect these practical results with the first chapter of Gilmore's 2008 text and you have a proof of Galois' work on the insolvability of the quintic by radicals. I never got it from crappy formal training overloaded with blinding abstraction. These cover the first two of the remaining blocks.

Is the quintic really unsolvable? Yes. By radical extensions. Galois showed us this. But mathematicians have built a mathematical technology, if you will, to get around the limitations of solvability by radicals. They have built on the theory of elliptic functions and elliptic integrals to deal with questions of solvability. I only know one elliptic integral well. It arises from the study of the pendulum for larger angles, where one can't use the Taylor approximation $\sin\theta \approx \theta$. Find impossibility. Find workaround. Be water my friend. Can't trisect an angle, double the cube, square the circle? Get the hell out of here! Go pick up something more than a straightedge and compass. C'mon. There are pathologies with vector (or gauge) potentials. Workaround. Connections over principle bundles. The best step-by-step book treating connections on bundles is "Geometry, Topology and Physics," M. Nakahara, Institute of Physics Publishing, 1990. The relevant material is in chapters 9 and 10, which I found I could jump into without reading prior chapters. The early chapters, however, are worth a read, especially on homology and homotopy. I understand there is a second edition. The book, "Topology, Geometry and Gauge Fields," by G. L. Naber, Springer 1997, I found impenetrable. Yet Naber does do a good job of connecting the mathematics to the physics at the word level, feeding my understanding. This material is drier—apologies.

Assuming that you have a junior level background in quantum mechanical addition of angular momentum, the final block on algebra is COMPLETELY SELF CONTAINED In FULL DETAIL. You will study Addition of Angular Momentum in terms of tensor products and direct products matrix methods, ladder operators and Young's tableaux. We will cover SU(2) for electron spin and Isospin and SU(3) for the quarks u, d and s. As a useful aside yielding selection rules, we will present in FULL DETAIL the application of the Wigner-Eckart theorem applied to scalars and vectors, and then extend the results to tensors. You may skip this if you're more interested in particle physics such as the Standard Model and its extensions to Grand Unified Theories (GUTs) and Theories of Everything (TOEs). What you will learn for SU(2) and SU(3) will be generalized in terms of the Cartan subalgebra with attendant ladder operators, root and weight diagrams and Dynkin diagrams so that the 2nd half of Jones is becomes much more readable, and so that you might have some chance of working through trash like Georgi's book on Lie algebras and particle physics.

Part IV. (The first six pages quickly review sophomore algebra.) We've just finished seeing the algebraic topology of continuous groups in terms of cosets and invariant (or normal) subgroups. Here are their discrete versions. Consider the symmetries of an equilateral triangle on the plane.



We may rotate by 0°, the identity operation e. We may rotate to the left by $c = 120°$. We may rotate left by $c^2 = 240°$. We may reflect the left diagonal b. We may reflect about the right diagonal b_2. We may reflect about the vertical diagonal b_3. These operations lead to finite, discrete group D_3 with table:

$*$	e	c	c^2	b	b_2	b_3
e	e	c	c^2	b	b_2	b_3
c	c	c^2	e	b_3	b	b_2
c^2	c^2	e	c	b_2	b_3	b
b	b	b_2	b_3	e	c	c_2
b_2	b_2	b_3	b	c_2	e	c
b_3	b_3	b	b_2	c	c_2	e

The element e is a (proper) subgroup of D_3, as are the elements $C_3 = \{e, c, c^2\}$. There are no other proper subgroups. We write $e < C_3 < D_3$, where D_3 is the dihedral group of the triangle, C_3 is the cyclic group of the triangle, and e is the trivial subgroup. (A cyclic group is any group generated by a single element; for the triangle we could pick c to be the generator as $c \cdot c = c^2$, $c^2 \cdot c = e$, $e \cdot c = c$.) The subgroup C_3 has the following one-dimensional representation

$$D(e) = 1, \qquad D(c) = e^{\frac{2\pi i}{3}}, \qquad D(c^2) = e^{\frac{4\pi i}{3}}.$$

These are 1×1 dimensional unitary matrices. Each of these matrices is irreducible (so, peeking ahead, $C_3 = D(e) \oplus D(c) \oplus D(c^2)$; note that $1^2 + 1^2 + 1^2 = 3$; this will have meaning).

Conjugacy. $a, b \in G$ are ***conjugate*** if there exists ($\exists$) $g \in G \ni a = gbg^{-1}$. We call g the conjugating element (fun) and we write $a \sim b$. (Back in continuous groups this is the commutator, a measuring device indicating how much the group deviates from being commutative (abelian).)

Conjugacy classes. $[a] = \{b | b \sim a\}$. This reads, the ***conjugacy class*** of a is the set of all b that are equivalent to a. (It's tacit that $a, b \in G$. The tilde denotes ***equivalence relation***.

For C_3 each element is a conjugacy class, as is the case for all abelian groups; $\forall\, a \in G$, $a = agg^{-1} = gag^{-1}$ for some $g \in G$. For D_3 we have three conjugacy classes: $[e]$, $[c, c^2]$, $[b, b_2, b_3]$. Certainly $e = geg^{-1}\ \forall\, g \in G$. Check all the cases, *i.e.*, $b = b_3 b_2 b_3^{-1} = b_3 b_2 b_3 = b_3 c = b_2$, and so on.

In terms of 2×2 matrices, the group D_3 elements are represented by

$$e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad c = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}, \qquad c^2 = \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix},$$

$$b = \frac{1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}, \qquad b_2 = \frac{1}{2}\begin{pmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}, \qquad b_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The subgroup C_3 is represented by the first three of these six matrices. The ***dimension of a representation*** is the dimension of the space on which it acts. We have seen both a one-dimensional representation of C_3 and a two-dimensional representation. When we define the regular representation, we will see that it is three-dimensional for C_3, with a concomitant geometric picture, and that for D_3 the regular representation is six-dimensional. The dimensionality of the regular representation will be the size of the group. (Note: This is sophomore/junior level modern algebra. The pace will be fast through these basics.)

Left (**right**) **cosets**. Given a subgroup $H \leq G$ the left cosets are $\{gH\}$ (or $\{Hg\}$ for right cosets).

Lagrange's Theorem—If H is a subset of G $(H \leq G)$, then $[H]|[G]$ (read the order of H divides the order of G. (The proof is sophomore modern algebra.)

Example 1—$[C_3]|\{D_3] = 3|6 = 2$. Three divides six equaling two. Another notation is $D_3/C_3 = 2$.

Normal subgroup. If $H \leq G$, we say H is a ***normal*** (***invariant***) ***subgroup*** of G $(H \trianglelefteq G)$ if $\forall \; g \in G, \; gHg^{-1} = H$.

Example 2—$C_3 \trianglelefteq D_3$. Let's check. I'll be abusive with notation to cut the work down. Pick any element from each of the three sets and will get $\{b, b_2, b_3\} \cdot \{e, c, c^2\} \cdot \{b, b_2, b_3\}^{-1} = \{e, c, c^2\} = C_3$. This is also true for $\{e, c, c^2\}\{e, c, c^2\}\{e, c, c^2\}^{-1} = \{e, c, c^2\} = C_3$.

There are two additional groups you should be aware of, the ***permutation group*** S_n and the ***alternating group*** A_n. Depending on the author, you will find the multiplication rule of these groups reversed—it's confusing. Here is an element of the permutation group S_3 on three symbols:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}.$$

The symbols 2 and 3 have been exchanged. An example multiplication of two members of S_3 is

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = (1\ 2).$$

The (1 2) notation is called cycle notation. I started on the right parenthesis, noting that 3 goes to 2, then in the left parenthesis, 2 goes to 3; so 3 remains unaffected. Then we go back to the right parenthesis to see that 2 goes to 1, then over to the left parenthesis to see that 1 goes to 1. So, effectively 2 goes to 1. Then lastly, from the right parenthesis I see that 1 goes to 3, over to the left parenthesis to see that 2 goes to 2. Thus 1 goes to 2. The shorthand cycle notation captures this group multiplication succinctly with (1 2); 1 goes to 2; 2 goes to 1; 3 remains unchanged. Given three symbols {1,2,3}, we know there are 3!=6 permutations, thus S_3 is a group of order 3. The alternating group on n symbols, A_n is the group of even permutations of S_n. A_n has $\frac{n!}{2}$ elements. Example of an odd permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 2 & 1 \end{pmatrix} = (1\ \ 5)(1\ \ 3)(2\ \ 4).$$

We cannot express the group element on the left by an even number of 2-cycles, *e.g.*, (1 5). Had the group element been expressible by an even number of 2-cylces it would have been an even permutation, and a member of A_5. We trouble about these because of the following theorem.

Cayley's Theorem—Every finite group is isomorphic to a subgroup of S_n. That is

$$g \to \prod(g) = \begin{pmatrix} 1 & 2 & \dots & n \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix}.$$

There is a sophomore level proof I'll level to an introductory book.

Quotient group (Factor group). Let me first give you an example using D_3. Let $E = C_3 = \{e, c, c^2\}$ and let $B = \{b, b_2, b_3\}$. Observe that E and B form a group:

$*$	E	B
E	E	B
B	B	E

The group is labeled D_3/C_3, which is a ***quotient*** or ***factor group***. This example jives with Lagrange's theorem for $|D_3/C_3| = 2$, the order of this quotient group is 2 = 6/3. D_3/C_3 is a mini version of D_3if you will thanks to equivalence relationships.

Homomorphism. A ***homomorphism*** is a mapping f from one set A to another set B preserving some structure in A.

Example 3—We may map the integers to the even integers by $n \to 2n$. The integers under addition form a group for they are closed under addition, have additive inverses, 0 is the identity, adding is associative. All this applies to the even integers. Our mapping is a homomorphism persevering the group structure of the integers under addition to the even integers.

Isomorphism. A homomorphism that is 1-1 and onto is an ***isomorphism***. In sophomore level mathematics you learn that you may map the integers to the rational numbers, but sets are countable. There is a 1-1 homomorphism linking the two sets—an isomorphism. Some people say injection instead of 1-1, and surjection for onto, and bijection for 1-1 and onto. I guess it sounds sexy.

Image. Given a mapping between two sets A and B by f, the ***image*** of f is $\mathrm{Im}f = \{b \epsilon B | b = f(a) \text{ for some } a \epsilon A\}$. Images may be 1-1, onto, or 1-1 and onto. **Preimages** are inverse images.

Kernel. The ***kernel*** $\mathrm{Ker}f = \{a \epsilon A \mid f(a) = e_B \epsilon B\}$, where e_B is the identity element of B.

Isomorphism theorem. If $f: G \to G$ is a homomorphism of G into G' with kernel K, then the image of f is isomorphic to the G/K. $\mathrm{Im}(f) \cong G/K$.

Think about this, if the kernel $K = e$, the identity element, the order of $\{e\}$ is one, and by Lagrange's theorem, $|G|/|e| = |G|$. So the mapping f is 1-1. G and G' are the same size and have the same algebra. They are different only in as much their symbology is different, $e.g.$, matrix representations versus analytic realizations.

Up to this point, we have the minimum algebra we need to proceed. If these definitions and theorems become fuzzy, review the triangle group D_3, or whip out the square and work out its group structure, then check the theorems with concrete examples. We begin the Jones notes. Think of group representations in terms of matrix representations.

A ***representation*** of dimension n of a group G is defined by $D: G \overset{\text{homo.}}{\longrightarrow} G\ell(n, \mathbb{C})$ where D are an $n \times n$ nonsingular complex matrices. IMPORTANT. Two $[n]$ representations $D^{(1)}, D^{(2)}$ of a group G are equivalent (~) if $D^{(1)}(g) = SD^{(2)}(g)S^{-1}\ \forall\ g \in G,\ S$ independent of g. This is a ***similarity transformation***

$$D^{(i)}(gg') = D^{(i)}(g)D^{(i)}(g').$$

Thus

$$SD^{(2)}(gg')S^{-1} = SD^{(2)}D^{(2)}(g')S^{-1} = SD^{(2)}(g)S^{-1}SD^{(2)}(g')S^{-1} = \left(SD^{(2)}S^{-1}\right)\left(SD^{(2)}(g')S^{-1}\right)$$

$$= D^{(1)}(g)D^{(1)}(g') = D^{(1)}(gg').$$

Example 4—In quantum mechanics one often seeks to diagonalize a matrix (change its basis to a diagonal basis) by similarity transformation. Consider

$$H = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

The trace of the matrix is 2. The determinant is 0. This gives us two constraints on the eigenvalues.

$$\lambda_1 + \lambda_2 + \lambda_3 = 2, \qquad \lambda_1\lambda_2\lambda_3 = 0.$$

Here is something that fits the constraints: $\lambda_1 = 2,\ \lambda_2 = 0,\ \lambda_3 = 0$. We could have proceeded with more traditional matrix algebra:

$$\det\begin{vmatrix} 1-\lambda & 0 & 1 \\ 0 & -\lambda & 0 \\ 1 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)(-\lambda(1-\lambda) - 0) + \lambda = 0.$$

The resulting polynomial is $\lambda^3 - 2\lambda^2 = \lambda^2(\lambda - 2) = 0$. The roots are $\{2,0,0\}$. For each root, we then find the eigenvalues as follows. Pick eigenvalue 2.

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = 2\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}.$$

Working this out means

$$\begin{pmatrix} x^1 + x^3 \\ 0 \\ x^1 + x^3 \end{pmatrix} = 2\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}.$$

This can only be so if $x^2 = 0$ and $x^1 = x^3$. Keeping it simple, let $x^1 = x^2 = 1$. Then indeed

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

The eigenvector is an eigenstates of the physical system. In quantum mechanics eigenstates are normalized to 1 because of its probabilistic interpretation. The dot product (or length) of the eigenstates is

$$(1\ 0\ 1)\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 2.$$

So let the normalized eigenvector be

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\end{pmatrix}.$$

Now the normalized length is 1. The remaining eigenvectors are

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix} \text{ and } \begin{pmatrix}0\\1\\0\end{pmatrix}.$$

The dot product of any two distinct normalized eigenvectors is 0. They are mutually orthogonal. From these eigenvectors, construct the matrix

$$U = \begin{pmatrix}\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}}\\ 0 & 1 & 0\\ \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}}\end{pmatrix}.$$

Then note that

$$H' = \begin{pmatrix}2 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{pmatrix} = U^{\dagger}HU,$$

where the dagger indicates complex conjugate transpose. Our matrices happen to be real-valued. The new eigenvectors are

$$\begin{pmatrix}1\\0\\0\end{pmatrix}, \quad \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad \begin{pmatrix}0\\0\\1\end{pmatrix},$$

which are the new basis vectors for the new coordinate system in which H has been transformed to the "diagonal" basis. H is equivalent to H' and conversely. The two are connected by a similarity transformation. The probabilistic underpinning of quantum mechanics forces the H to be hermitian matrices and the U to be unitary matrices.

The ***character*** of a representation D of a group G is $\chi = \{\chi(g) \mid g \in G\}$, where $\chi(g)$ is the trace of the matrix representing the element g (the sum of the diagonal elements). Simple theorems tell us that $\mathrm{Tr}(ABC) = \mathrm{Tr}(ACB) = \cdots$ and the remaining four permutations of the symbols. Also, $\mathrm{Tr}(SD(g)S^{-1}) = \mathrm{Tr}\big(SS^{-1}D(g)\big) = \mathrm{Tr}\big(D(g)\big)$.

Example (5)—The theorems ahead are dry enough and abstract enough that I lead the theory with an example of what they will empower us to do, namely to break a group down into its irreducible representations. You might have to iterate more than once between this example and the next example, as well as the theorems to get the idea; at least this was the case for me over an extended period of studies interspersed with breaks. Enjoy the next two pages written for now without justification.

Recall C_3:

$*$	e	c	c^2
e	e	c	c^2
c	c	c^2	e
c^2	c^2	e	c

One theorem will tell us the number of irreducible representations equals the number of conjugacy classes. This is three for C_3. One of these conjugacy classes, as always, is the identity e. Is c conjugate to c^2? Yes if there is an element $g \in C_3 \ni \ c = gc^2g^{-1}$. Well $c \neq c^2c^2c^{-2} = c^2c^2c = c^2e = c^2$, but $c = e^2ce^{-2} = c$. So $c \sim c$, a conjugacy class of size 1; ditto for c^2. Another theorem ahead will state that the sum of the squares of the dimensionalities of the irreducible representations must sum to the order (size) of the group: $n_1^2 + n_2^2 + n_3^2 = 3$, implying C_3 is the direct sum of three one-dimensional representations (only $1^2 + 1^2 + 1^2 = 3$). This is stupidly obvious for C_3, but as you'll see for D_3 we have $1^2 + 1^2 + 2^2 = 6$, two one-dimensional representations and one two-dimensional representation.

The characters of a one-dimensional representation must mimic the group multiplication. In particular $\chi(c^2) = \left(\chi(c)\right)^2$ and $\chi(c)^3 = \chi(c^3) = \chi(e) = 1$; this can be so if $\chi(c)$ is a root of unity: 1, $\omega = e^{\frac{2\pi i}{3}}$, or $\omega^2 = e^{\frac{4\pi i}{4}}$. So the character table begins to look like

C_3	e	c	c^2
$D^{(1)}$	1	1	1
$D^{(2)}$	1	ω	ω^2
$D^{(3)}$	1	ω^2	ω

If $\chi^{(i)}, i = 1{,}2.3$ are the three rows above (three row vectors), another result will be that the $\chi^{(i)}$ divided by the order of the group are orthogonal. Let's check some case using the scalar product:

$$\langle \chi^{(1)}, \chi^{(1)} \rangle = \frac{1}{3}(1 + 1 + 1) = 1,$$

$$\langle \chi^{(1)}, \chi^{(2)} \rangle = \frac{1}{3}(1 + \omega + \omega^2) = 0,$$

by virtue of the factorization of $\omega^3 - 1 = (\omega - 1)(1 + \omega + \omega^2)$. Consider the (non-regular representation) 3×3 rotation matrices about the z-axis as a representation of C_3:

$$e = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}_{\theta=2\pi/3}, \quad c^2 = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}_{\theta=4\pi/3}.$$

Let's define the ***compound character*** χ. For C_3 (with matrices above) $\chi = (\mathrm{Tr}(e) = 3, \mathrm{Tr}(c) = 0, \mathrm{Tr}(c^2) = 0)$; is this so for other representations? If so, when? We'll see. In χ we've constructed a three dimensional object with three orthogonal unit vector components of "simple" traces:

$$\chi = \left(\chi(e), \chi(c), \chi(c^2)\right) = (3{,}0{,}0) = a_1\chi^{(1)} + a_2\chi_2^{(2)} + a_3\chi^{(3)},$$

where $a_\nu =< \chi, \chi^\nu >,\ \nu = 1,2,3$. For example, $a_1 = \frac{1}{3} < \chi, \chi^1 >= \frac{1}{3} 3 = 1$. This is very quantum mechanical formalism in the sense of Dirac's formalism of quantum mechanics.

The group $C_3 = D^{(1)} \oplus D^{(2)} \oplus D^{(3)}$, a ***direct sum*** of three inequivalent one-dimensional irreducible representations with geometric interpretation: $D^{(1)}$ represents rotations about the z axis leading to $x' = x\cos\theta - y\sin\theta$ and $y' = y\cos\theta + x\sin\theta$. Now consider the sums

$$x' \pm y' = x(\cos\theta \pm i\sin\theta) \pm iy(\cos\theta \pm i\sin\theta) = (x \pm iy)\, e^{i\theta}.$$

For $D^{(2)}$ the angle $\theta = \frac{2\pi}{3}, \therefore e^{i\theta} = \omega$. Thus $x + iy$ is the geometric basis of $D^{(2)}$. With $\theta = \frac{4\pi}{3}$, we have $e^{i\theta} = \omega^2$. The representative of $D^{(3)}$ is $x - iy$. These two are one-dimensional lines on the complex plane—the diagonals of our equilateral triangle. The one-dimensional z-axis pokes out perpendicularly to the lines $x \pm iy$. The diagonalized "Fourier" decomposition of the group C_3 is

$$\begin{pmatrix} D^{(1)} = 1 & 0 & 0 \\ 0 & D^{(2)} = e^{\frac{2\pi i}{3}} & 0 \\ 0 & 0 & D^{(3)} = e^{\frac{4\pi i}{3}} \end{pmatrix} = D^{(1)} \oplus D^{(2)} \oplus D^{(3)}.$$

Each representation acts on $\mathbb{R}^3$ triples (x, y, z) as we have described. For C_3 each representation is a one-dimensional unitary matrix—a complex number. For the quantum physicist we now know the group C_3 acts on three orthogonal kets, $|1\rangle, |2\rangle$ and $|3\rangle$.

A glaring question at this point might be why I didn't use the two-dimensional matrix representation to compute χ instead of the three-dimensional rotation matrices? Could I have used 4×4 (or larger) square matrix representations? The theorems and constructions themselves lead to matrices of the proper size for our traces. When you are plowing thru the theorems, recall this example as the particular case being generalized. Character tables underlie the insolvability of the quintic. A word of advice before digging in. Plug in examples from the C_3 and D_3 groups using their various representations every chance you get. Here is the ***regular representation*** for C_3:

$$D(c)=\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix},\qquad D(e)=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix},\ D(c^2)=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}.$$

Here's how to make these matrices using C_3's group table and orthogonal bra-kets. Starting with C_3 let

$$D(g_1=e)=\begin{pmatrix}\langle g_1|g_1|g_1\rangle & \langle g_1|g_1|g_2\rangle & \langle g_1|g_1|g_3\rangle\\ \langle g_2|g_1|g_1\rangle & \langle g_2|g_1|g_2\rangle & \langle g_2|g_1|g_3\rangle\\ \langle g_3|g_1|g_1\rangle & \langle g_3|g_1|g_2\rangle & \langle g_3|g_1|g_3\rangle\end{pmatrix}=\begin{pmatrix}\langle e|e|e\rangle & \langle e|e|c\rangle & \langle e|e|c^2\rangle\\ \langle c|e|e\rangle & \langle c|e|c\rangle & \langle c|e|c^2\rangle\\ \langle c^2|e|e\rangle & \langle c^2|e|c\rangle & \langle c^2|e|c^2\rangle\end{pmatrix}$$

$$=\begin{pmatrix}\langle e|e\rangle & \langle e|c\rangle & \langle e|c^2\rangle\\ \langle c|e\rangle & \langle c|c\rangle & \langle c|c^2\rangle\\ \langle c^2|e\rangle & \langle c^2|c\rangle & \langle c^2|c^2\rangle\end{pmatrix}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$$

where in our notation two group elements g_i, g_j are parallel if $i=j$ (that is $\langle g_i|g_j\rangle=1$ if $i=j$) otherwise if $i\neq j$ the group elements are orthogonal (that is $\langle g_i|g_j\rangle=0\ if\ i\neq j$). Proceeding to c we get the following 3×3 matrix:

$$D(g_2=c)=\begin{pmatrix}\langle g_1|g_2|g_1\rangle & \langle g_1|g_2|g_2\rangle & \langle g_1|g_2|g_3\rangle\\ \langle g_2|g_2|g_1\rangle & \langle g_2|g_2|g_2\rangle & \langle g_2|g_2|g_3\rangle\\ \langle g_3|g_2|g_1\rangle & \langle g_3|g_2|g_2\rangle & \langle g_3|g_2|g_3\rangle\end{pmatrix}=\begin{pmatrix}\langle e|c|e\rangle & \langle e|c|c\rangle & \langle e|c|c^2\rangle\\ \langle c|c|e\rangle & \langle c|c|c\rangle & \langle c|c|c^2\rangle\\ \langle c^2|c|e\rangle & \langle c^2|c|c\rangle & \langle c^2|c|c^2\rangle\end{pmatrix}$$

$$=\begin{pmatrix}\langle e|c\rangle & \langle e|c^2\rangle & \langle e|e\rangle\\ \langle c|c\rangle & \langle c|c^2\rangle & \langle c|e\rangle\\ \langle c^2|c\rangle & \langle c^2|c^2\rangle & \langle c^2|e\rangle\end{pmatrix}=\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix}.$$

The 3×3 matrix $D(g_3=c^2)$ follows from this notation. This is how we build the regular representation for finite groups.

Reducibility. A representation of dimension $m+n$ is ***reducible*** if $D(g)$ takes the form

$$D(g)=\begin{pmatrix}A(g) & C(g)\\ 0 & B(g)\end{pmatrix}\ \forall\ g\ \epsilon\ G,$$

A, B, C, being matrices of, respectively, dimension $m \times m,\ m \times n,\ n \times n$. The matrices are closed:

$$D(g)D'(g) = \begin{pmatrix} A(g) & C(g) \\ 0 & B(g) \end{pmatrix}\begin{pmatrix} A'(g) & C'(g) \\ 0 & B'(g) \end{pmatrix} = \begin{pmatrix} A(g)A'(g) & AC' + CB' \\ 0 & B(g)B'(g) \end{pmatrix} = \begin{pmatrix} A(gg') & C(gg') \\ 0 & B(gg') \end{pmatrix},$$

where $C(gg') = AC' + CB'$. This is certainly true of the 3×3 rotation matrix around the z-axis if you think of it in the form

$$R = \left(\begin{array}{cc|c} * & * & \\ * & * & \\ \hline & & 1 \end{array}\right).$$

The representation is ***completely reducible*** if $C(g) = 0$. Then $D(g)$ is the ***direct sum*** $D(g) = A(g) \oplus B(g)$. If A and B are reducible, we proceed until we reach an ***irreducible representation***.

A set of vectors $\{\boldsymbol{e}_i\}, i = 1,2, \dots, m$ is ***linearly independent*** if there is a non-trivial combination which yields the null vector $\sum_{i=1}^{m} \lambda_i \mathbf{e}_i = 0 \implies \lambda_i = 0\ \forall i$. A linearly independent set of vectors $\{\mathbf{e}_i\}, i = 1,2, \dots, m$ forms a ***basis*** of a vector space V if they span the space; any $u \epsilon U$ can be expressed as a linear combination of the $\mathbf{e}_i$, $\mathbf{u} = \sum_{i=1}^{m} u_i \mathbf{e}_i$. Recall what ***a vector space*** is. If we adjoin a product $\boxdot$ with the right properties we get an algebra. Recall that for real, antisymmetric matrices are product was the commutator, making the algebra a Lie algebra. There are many more products we can think of in search of useful products—you'll see. Now for a little undergraduate linear algebra.

A ***linear transformation*** on a vector space V is a map $T: V \to V$ such that $T(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v})$, where $\mathbf{u}, \mathbf{v}$ are vectors in V. Relating this to matrices, given a basis $\{\mathbf{e}_i\}$ of V, the map T has a concrete representation as a matrix D_{ij}. This is defined by giving the transforms of each of the basis vectors $\mathbf{e}_j$. Because $\{\mathbf{e}_i\}$ form a basis, each such transform is a linear combination of the basis vectors, which we write in the form: $T\mathbf{e}_j = D_{\mathrm{ij}}\mathbf{e}_{j.}$ We are using the Einstein summation convention whereby a summation over a repeated index is understood. Knowing the transforms of all of the basis vectors we are able to write down the transform of a general vectors $\mathbf{u} = u_i \mathbf{e}_i$, namely

$$T\mathbf{u} = T(u_j\mathbf{e}_j) = u_j(T\mathbf{e}_j) = u_j D_{\text{ij}}\mathbf{e}_i.$$

So, writing $T\mathbf{u} = \mathbf{v} = v_i\mathbf{e}_i$, we can identify $v_i = D_{ij}u_j$.

Similarity. The matrix D was derived from T by reference to a particular basis $\{\mathbf{e}\}$ which (as we have seen by similarity transformations) is not unique (!). Let's review changing basis from $\{\mathbf{e}_i\}$ to $\{\mathbf{f}_i\}$. We will obtain a different matrix D'. We define this by:

$$T\mathbf{f}_i = D'_{ij}\mathbf{f}_i,$$

and the new coordinates of $\mathbf{u}$ defined by writing $\mathbf{u} = u'_j\mathbf{f}'_i$ will be transformed to

$$v'_i = D'_{ij}u'_j.$$

In matrix form this is $v' = D'u'$. On the other hand we may relate u' to u and v' to v by expressing each old basis vector $\mathbf{e}_i$ as some linear superposition of the new basis vectors $\{\mathbf{f}_j\}$:

$$\mathbf{e}_i = S_{ji}\mathbf{f}_j.$$

Thus

$$\mathbf{u} = u_i\mathbf{e}_i = u_i S_{ji}\mathbf{f}_j,$$

so that

$$u'_j = u_i S_{ji},$$

or in matrix form, $u' = Su$. (Under a change of basis, do you see the contravariant transformation with S_{ji} appearing to the left of the basis vectors, but sitting to the right of the contravariant vector components?) Similarly $v' = Sv$. Altogether then, $v' = Sv = S(Du) = SD(S^{-1}u')$. Clearly $D' = SDS^{-1}$.

G-Module. If V is a vector space, G is a group ∋ $\forall\, g \in G,\ T(g'g) = T(g')T(g)$, then G is a ***G-module***. When we were working with continuous symmetries, G was the topological (continuous) group $\mathcal{T}$ while G was the geometric manifold.

Reducibility updated. In the language of G-modules and matrices, reducibility corresponds to the existence of a closed submodule: $\mathbf{u} \in U \subset V \implies T(g)\mathbf{u} \in U$. Let $\{\mathbf{e_i}\}$ be a basis for $U, i = 1,2, \dots, m$. We can extend this basis by $\{\mathbf{e_i}\}$, $i = m+1, m+2, \dots m+n$ to form a basis of the larger space V of dimension $m+n$.

Relative to this basis, the matrix $D(g)$ corresponding to the linear transformation $T(g)$ is given by $T(g)\mathbf{e}_j = D_{ij}(g)\mathbf{e}_i$. The closure of U is reflected by $D_{ij} = 0$ when $j = 1,2, \dots, m$ when $i = m + 1, \dots, m+n$. That is, updating our matrix version of reducibility:

$$D(g) = \begin{pmatrix} A_{m\times m}(g) & C_{m\times n}(g) \\ 0_{m\times n} & B_{n\times n}(g) \end{pmatrix}.$$

For a finite group (and a compact group (later theorem)) $C(g)$ can be set to zero. To prove this it is sufficient to show that all representations of a finite group are equivalent to unitary transformations.

Scalar (***dot***) ***product***. If V is a vector space with $\mathbf{u}, \mathbf{v} \in V\ \exists\ c \in \mathbb{C}$ and $c = (\mathbf{u}, \mathbf{v})$,

$$(\mathbf{u}, \mathbf{v}\,) = (\mathbf{v}, \mathbf{u})^* \ \ \text{(hermitian)}$$

$$(\mathbf{u}, \mathbf{u}\,) = 0 \ \ \text{(positive definite)}$$

$$(\mathbf{w}, \alpha\mathbf{u} + \beta\mathbf{v}) = \alpha(\mathbf{w}, \mathbf{u}) + \beta(\mathbf{w}, \mathbf{v}) \ \ \text{(linearity)}$$

$$\|\mathbf{u}\| = \sqrt{(\mathbf{u}, \mathbf{u})} \ \ \text{(norm)},$$

then $(\mathbf{u}, \mathbf{v})$ is the scalar (or dot) product.

Orthonormal basis (ONB). Given any basis $\{\mathrm{f}_i\}$ of a vector space V, it is possible to construct an ONB $\{\mathbf{e}_i\} \ni (\mathbf{e}_i, \mathbf{e}_j) = \delta_{ij}$. The procedure is known as Gram-Schmidt orthogonalization:

$$\mathbf{e}_1 = \mathbf{f}_1/\|\mathbf{f}_2\|$$

$$\mathbf{e}_2 = \mathbf{f}_2 - \frac{\mathbf{e}_1, \mathbf{f}_2}{\|\mathbf{e}_1, \mathbf{f}_2\|}$$

$$\mathbf{e}_3 = \mathbf{f}_3 - \frac{\mathbf{e}_1, \mathbf{f}_3}{\|\mathbf{e}_1, \mathbf{f}_3\|} - \frac{\mathbf{e}_2, \mathbf{f}_3}{\|\mathbf{e}_2, \mathbf{f}_3\|}, \ldots$$

Unitary transformations. A linear transformation T acting in V is unitary if $(T\mathbf{u}, T\mathbf{v}) = (\mathbf{u}, \mathbf{v})\ \forall\ \mathbf{u}, \mathbf{v} \in V$. The $T(g)$ are necessarily invertible if the transformation is unitary. Equivalently, we can write

$$(T\mathbf{u}, \mathbf{v}) = (\mathbf{u}, T^{-1}\mathbf{v}),$$

and equivalently given D_{ij} relative to some ONB $\{\mathbf{e}_i\}$, we necessarily have $D_{ij}^{\dagger} = (D^{-1})_{ji}$. For unitary matrices $D^{\dagger} = (D^*)^t = D^{-1}$; the complex, conjugate equals the inverse. For a hermitian matrix, $D^{\dagger} = D$.

Complete reducibility in terms of an ONB. Now that we have a scalar product, we may choose the basis of V to be an ONB. Any vector $\mathbf{w}$ in the space W spanned by the additional vectors $\{\boldsymbol{e}_i\}, i = m+1, \ldots m+n$ is orthogonal to any $\boldsymbol{u}$ in U:

$$W = \{\mathbf{w}\ \in V \mid (\mathbf{w}, \mathbf{u}) = \ 0\ \ \forall\ \mathbf{u} \in U\}.$$

In a coordinate-free language, complete reducibility corresponds to the closure of the subspaces U and V under scalar product. This is always true if the $T(g)$ are unitary transformations wrt to the scalar product, for then

$$(T(g)\mathbf{w}, \mathbf{u}) = (\mathbf{w}, T^{-1}(g)\mathbf{u}).$$

The closure of U then guarantees that $T^{-1}(g)\mathbf{u} = T(g^{-1})\mathbf{u} = \mathbf{u}' \,\epsilon\, U$. Hence $(T(g)\mathbf{w}, \mathbf{u}) = (\mathbf{w}, \mathbf{u}') = 0$. Cleary then $T(g)\mathbf{w} = \mathbf{w}' \,\epsilon\, W$.

Maschke's Theorem. All reducible representations of a finite group are completely reducible—decomposable into a direct sum of irreducible representations. (Hint: Use (cleverly create) a scalar product which always gives us a unitary representation.)

Proof. Define the ***group invariant scalar product*** by building up from the original scalar product, but now sum over all group elements:

$$\{\mathbf{v}, \mathbf{v}'\} = \frac{1}{[g]} \sum_{g \epsilon G} (T(g)\mathbf{v}, T(g)\mathbf{v}').$$

The $\{\mathbf{v}, \mathbf{v}'\}$ are generic members of $V = U \oplus W$. Consider $\{T(h)\mathbf{v}, T(h)\mathbf{v}'\}$, $h \,\epsilon\, G$. The group invariant scalar product becomes:

$$\begin{aligned} &\{T(h)\mathbf{v}, T(h)\mathbf{v}'\} \\ &\quad = \frac{1}{[g]} \sum_{g \epsilon G} (T(g)T(h)\mathbf{v}, T(g)T(h)\mathbf{v}') \\ &\quad = \frac{1}{[g]} \sum_{g \epsilon G} (T(gh)\mathbf{v}, T(gh)\mathbf{v}') = \frac{1}{[g]} \sum_{g \epsilon G} (T(g')\mathbf{v}, T(g')\mathbf{v}'). \end{aligned}$$

Just as was true for group invariant left (right) measures for continuous groups, $\sum_g = \sum_g'$ for discrete groups. Thus we have

$$\{T(h)\mathbf{v}, T(h)\mathbf{v}'\} = \{\mathbf{v}, \mathbf{v}'\}.$$

Evidently, the $T(h)$ are unitary wrt to the group invariant scalar product which was cleverly created to do exactly this. In terms of matrices, the $T(h)$ will be realized as a reducible unitary matrix $D'(h)$ if we choose a basis $\{\mathrm{f}_i\}$ orthonormal wrt the group invariant scalar product, whose first members m span U. The $D'(h)$ will be related by the similarity transformation $D'(h) = SD(h)S^{-1}$, where S is the matrix effecting the change of basis from $\{\mathrm{f}_i\}$ to $\{\mathbf{e}_i\}$. (The representation D is equivalent to the reducible unitary representation D', which is completely reducible.) In pictures:

$$V = \begin{pmatrix} \underline{U|} & 0 \\ 0 & \overline{|W} \end{pmatrix}.$$

What does this group invariant scalar product look like for, say, C_3? Let $\mathbf{v} = (a, b, d)$ and $\mathbf{v}' = (e, f, g)$ be vectors in $\mathbb{R}^3$. The usual scalar product is

$$(a\ b\ \ d) \cdot (e\ f\ g)^T = (a\ b\ c) \cdot \begin{pmatrix} e \\ f \\ g \end{pmatrix} = ae + bf + cg.$$

With our three-dimensional rotation matrix representation, the group invariant scalar product looks like

$$\frac{1}{[g]} \sum_{g \epsilon G} (T(g)\mathbf{v}, T(g)\mathbf{v}')$$

$$= \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{v} \cdot \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{v}'\right]^T + \frac{1}{3}\begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{v}$$

$$\cdot \left[\begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{v}'\right]^T + \frac{1}{3}\begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{v} \cdot \left[\begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{v}'\right]^T$$

Schur's first lemma. Any matrix which commutes with all the matrices of an irreducible representation is a multiple of the identity,

$$BD(g) = D(g)B\ \forall\ g \epsilon G \implies B = \lambda\mathbb{I}.$$

(Think in terms of the three matrices above.) In terms of linear operators,

$$\hat{B}T(g) = T(g)\hat{B} \; \forall \; g \in G \;\; \Longrightarrow \;\; \hat{B} = \lambda E,$$

where $\hat{B}: U \to U$ is the linear operator version of B, and E is the identity map $\mathbf{u} \to \mathbf{u}$.

Proof. Let's work with the linear operator version. Let $\mathbf{b}$ be an eigenvector of $\hat{B}$ with eigenvalueλ: $\hat{B}\mathbf{b} = \lambda\mathbf{b}$. Then

$$\hat{B}(T(g)\mathbf{b}) = T(g)\hat{B}\mathbf{b} = T(g)\lambda\mathbf{b} = \lambda\big(T(g)\big).$$

Then $T(g)$ is an eigenvector of $\hat{B}$ with the same eigenvalue λ. The space of all such eigenvectors is a vector space which is a subset of the space U. It's a G-module, being closed under the linear operators $\{T(g)\}$.

By irreducibility, the only submodules are the space U or the identity $\mathbf{0}$. Since every linear operator $\hat{B}$ has at least one proper eigenvector, $\hat{B}$ is not mapping vectors to the identity. In the complex numbers, $\det(B - \lambda\mathbb{I}) = 0$ has at least one root λ corresponding to an eigenvector $\mathbf{b}$. Thus the space of eigenvectors of $\hat{B}$ with eigenvalue λ is the whole space. So $\hat{B}\mathbf{u} = \lambda\mathbf{u} \;\; \forall \; \mathbf{u} \; \in U$ and $\hat{B} = \lambda E$ as claimed.

Schur's second lemma. Consider two inequivalent representations D and D'. In matrix form, Schur's second lemma states

$$BD(g) = D'(g)B \;\; \forall \; g \;\in G \;\; \Longrightarrow B = \hat{0}.$$

The corresponding linear operator $\hat{B}$ will be a map $\hat{B}: U \to U'$ between two vector spaces, each with two different representations. In coordinate-free language, the equation above is

$$\hat{B}T(g) = T'(g)\hat{B} \;\; \forall \; g \;\in G \;\; \Longrightarrow \hat{B} = \hat{0},$$

where $\hat{O}$ is the linear operator mapping every vector in U onto the null vector in U' ($\hat{O}\mathbf{u} = 0' \ \forall \, \mathbf{u} \in U$).

Before we get to the proof, what is this lemma telling us? We have 2×2 two-dimensional representations of C_3 (and D_3) and we have 3×3 three-dimensional representations of these groups. They are inequivalent representations. The only matrix that makes 2×2 commute with 3×3matrices is the $\hat{O}$ operator above. This remains true for two inequivalent representations even if they have the same dimensionality. Recall the one-dimensional irreducible representations of C_3, namely $C_3 = D^{(1)} \oplus D^{(2)} \oplus D^{(3)}$. There is no nonzero (one-dimensional) matrix B so that, say $B \cdot e^{\frac{2\pi i}{3}} = e^{\frac{4\pi i}{6}} \cdot B$.

Proof. Case (1). Suppose $\dim U = n < \dim U' = n'$. Consider the action of $T'(g)\hat{B}$ on an arbitrary vector $\mathbf{u}$ of U. From the above equation

$$T'(g)\hat{B}\mathbf{u} = \hat{B}T(g)\mathbf{u}.$$

Since U is a G-module, $T(g)\mathbf{u} \in U$. Thus,

$$T'(g)(\hat{B}\mathbf{u}) \in \hat{B}U$$

which means that $\hat{B}U = \mathrm{Im}\,\hat{B}$ is a submodule of U'. But U' is, by supposition, irreducible. So $\widehat{BU}$ must be either the whole space U' or the null vector $\mathbf{0}^1$. We can exclude the first case because we assumed from the start that $n < n'$. As the image of U the dimension of m of $\hat{B}U$ can't exceed that of U. So $m \leq n < n'$. We're left with $\hat{B} = \hat{O}$.

Case (2). If $n > n'$ then the kernel of the mapping is

$$K = \{\mathbf{k} \in U \mid \hat{B}\mathbf{k} = \mathbf{0}'\}$$

The kernel is a submodule of U since $\hat{B}\big(T(g)\big)\mathbf{k}) = T'(g)\hat{B}\mathbf{k} = \mathbf{0}'$. By the irreducibility of U, K is either the whole space U or it is the null vector $\mathbf{0}$. K must be nontrivial because the images of $\hat{B}\mathbf{e}_i$ of the n vectors of a basis of U cannot all be linearly independent since the mapping is reducing the dimensionality from n to the smaller number n'. Thus every vector $\mathbf{u} \epsilon U$must satisfy $\hat{B}\mathbf{u} = \mathbf{0}'$.

Case (3). Let $n = n'$. Again the kernel K is a submodule of U, but it can't be the null vector $\mathbf{0}$ because D and D' are inequivalent. Recall what I said about the situation when the kernel is the identity element, namely this means that the mapping is 1-1. We just assumed that D and D' are inequivalent. We conclude that $\hat{B} = \hat{O}$.

Physicist's aside—at this point Georgi connects Schur's lemmas with quantum physics. Georgi's version of Maschke's theorem is that every representation of a finite group is completely reducible. His proof involves the use of a projector operator P such that $PD(g)P = D(g)P\ \forall\ g \in G$. (This is the condition that P be an invariant subspace.) Georgi presses to a combined version of Schur's lemma: if $D_1(g)A = AD_2(g) \forall\ g \in G$ where D_1 and D_2 are inequivalent irreducible representations, then $A = 0$. Then Georgi shows that Schur's lemmas have strong consequences for the matrix elements of any quantum mechanical operator O that corresponds to an observable that is invariant under the symmetry transformations. This is because the matrix elements $\langle a, k, x|O|b. k. y\rangle$ behave like the A in Schur's lemma. The details of this connection are shown from the middle of page 14 to the middle of page 17 (2nd ed.). The rest of Georgi's chapter 1 is in parallel with the Jones-based notes presented here on finite groups, but using Dirac bra-ket notation and projection operators. **End physicist's aside**.

Fundamental orthogonality theorem—with examples. Let U_ν and U_μ be two G-modules carrying inequivalent irreducible representations of some given group G. (The indices may be continuous for compact groups, but let's let them be integers.) Pick an arbitrary linear mapping $\hat{A}$ from U_ν to U_μ and construct (yes we construct) the following operator

$$\hat{B} = \sum_g T^{(\mu)}(g)\hat{A}T^{(\nu)}(g^{-1}).$$

This involves the same (scalar product) sum that was constructed to make $T^{(\mu)}(g)$ unitary (under this constructed scalar product). (For compact groups we replace the sum with $\int d\mu(g)$; review Part III.)

Let h be any member of G and consider

$$T^{(\mu)}(h)\hat{B} = \sum_g T^{(\mu)}(h)T^{(\mu)}(g)\hat{A}T^{(\nu)}(g^{-1}).$$

Notice that this involves the same sum over group elements used in Maschke's theorem. Thus,

$$T^{(\mu)}(h)\hat{B} = \sum_g T^{(\mu)}\,(hg)\hat{A}T^{(\nu)}(g^{-1})$$

from the group property of the $T^{(\mu)}$. Let hg be g'. Then the argument of $T^{(\nu)}$ becomes $g^{-1} = g'^{-1}h$, and we know the sum over g is the same as the scrambled sum over g'. Thus

$$T^{(\mu)}(h)\hat{B} = \sum_g T^{(\mu)}(g')\hat{A}T^{(\nu)}(g'^{-1}h) = \sum_g T^{(\mu)}(g')\hat{A}T^{(\nu)}(h)$$

from the group property of the $T^{(\nu)}$. Notice the first three factors, summed over g' give $\hat{B}$ again. Thus

$$T^{(\mu)}(h)\hat{B} = BT^{(\nu)}(h).$$

Evidently $\hat{B}$ satisfies the conditions of Schur's second lemma; thus $\hat{V} = \hat{O}$ unless $\mu = \nu$. In the latter case the two representatives are the same. Well then Schur's first lemma applies, yielding $\hat{B} = \lambda E$. This can be expressed in matrix form as

$$\sum_g D_{ij}^{(\mu)}(g)D^{(\nu)}(g^{-1}) = \lambda_A^{(\mu)}\delta_{\mu\nu}\mathbb{I}.$$

The constant λ depends both on the label of the irreducible representation and on the choice of matrix A, which remains unspecified. Let's make life easy by specifying all elements of A to be zero except for $A_{rs} = 1$. So now write $\lambda_A^{(\mu)}$ as λ_{rs}^{μ}. Then the ij matrix element of the sum above reads

$$\sum_g D_{ir}^{(\mu)}(g) D_{sj}^{(\nu)}(g^{-1}) = \lambda_{rs}^{\mu} \delta^{\mu\nu} \delta_{ij}.$$

The $\lambda_{rs}^{(\mu)}$ can be found by setting $\nu = \mu$ and tracing (contracting with δ_{ij} summation that is). This gives

$$\sum_g \left(D^{(\mu)}(g^{-1}) D^{(\mu)}(g) \right)_{sr} = n_\mu \lambda_{rs}^{(\mu)}$$

where n_μ is the dimensionality of $D^{(\mu)}$. The matrix $D^{(\mu)}(g^{-1}) D^{(\mu)}(g)\ \forall\ g$ is just the unit matrix with matrix elements δ_{rs}. Thus $[g]\delta_{rs} = n_\mu \lambda_{rs}^{(\mu)}$. Substituting for $\lambda_{rs}^{(\mu)}$ in $\sum_g D_{ij}^{(\mu)}(g) D^{(\nu)}(g^{-1}) = \lambda_A^{(\mu)} \delta_{\mu\nu} \mathbb{I}$, we obtain the fundamental orthogonality relation for the matrices of irreducible representations as the equation below:

$$\sum_f D_{ir}^{(\mu)}(g) D_{sj}^{(\nu)}(g^{-1}) = \frac{[g]}{n_\mu} \delta^{\mu\nu} \delta_{ij} \delta_{rs}.$$

What does $\sum_g D_{ir}^{(\mu)}(g) D_{js}^{(\nu)}(g) = \frac{[g]}{n_\mu} \delta^{\mu\nu}\ \delta_{ij} \delta_{rs}$ look like? Is the result independent of the matrix representation? Consider for example the two-dimensional 2×2 representations of the dihedral group D_3:

$$e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad c = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}, \qquad c^2 = \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix},$$

$$b = \frac{1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix}, \qquad b_2 = \frac{1}{2}\begin{pmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}, \qquad b_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Explicitly (with $\mu = \nu$ so tha $\delta^{\mu\nu} = 1$ the LHS of the sum is

$$\sum_g D_{ir}^{(\mu)}(g)D_{js}^{(\nu)}(g^{-1}) = D_{ir}^{(\mu)}(e)D_{sj}^{(\mu)}(e^{-1}) + D_{ir}^{(\mu)}(c)D_{sj}^{(\mu)}(c^{-1}) + D_{ir}^{(\mu)}(c^2)D_{sj}^{(\mu)}\left(c^{2^{-1}}\right)$$

$$+ D_{ir}^{(\mu)}(b)D_{sj}^{(\mu)}(b^{-1}) + D_{ir}^{(\mu)}(b_2)D_{sj}^{(\mu)}(b_2^{-1}) + D_{ir}^{(\mu)}(b_3)D_{sj}^{(\mu)}b_3^{-1}).$$

Let, say, $i = j = r = s = 1$. Then

$$\sum_g D_{ir}^{(\mu)}(g)D_{js}^{(\nu)}(g) = D_{11}^{(\mu)}(e)D_{11}^{(\mu)}(e^{-1}) + D_{11}^{(\mu)}(c)D_{11}^{(\mu)}(c^{-1}) + D_{11}^{(\mu)}(c^2)D_{11}^{(\mu)}\left(c^{2^{-1}}\right)$$

$$+ D_{11}^{(\mu)}(b)D_{11}^{(\mu)}(b^{-1}) + D_{11}^{(\mu)}(b_2)D_{11}^{(\mu)}(b_2^{-1}) + D_{11}^{(\mu)}(b_3)D_{11}^{(\mu)}b_3^{-1})$$

$$= 1 + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + 1 = 3 = \text{RHS} = \frac{[g]}{n_\mu} = \frac{6}{2}\delta^{11}\delta_{11}\delta_{11}.$$

What about $i = j = r = 1;\ s = 2$? Then

$$\sum_g D_{ir}^{(\mu)}(g)D_{js}^{(\nu)}(g) = D_{11}^{(\mu)}(e)D_{12}^{(\mu)}(e^{-1}) + D_{11}^{(\mu)}(c)D_{12}^{(\mu)}(c^{-1}) + D_{11}^{(\mu)}(c^2)D_{12}^{(\mu)}\left(c^{2^{-1}}\right)$$

$$+ D_{11}^{(\mu)}(b)D_{12}^{(\mu)}(b^{-1}) + D_{11}^{(\mu)}(b_2)D_{12}^{(\mu)}(b_2^{-1}) + D_{11}^{(\mu)}(b_3)D_{12}^{(\mu)}b_3^{-1})$$

$$= 1 \cdot 0 - \frac{1}{2}\cdot\frac{\sqrt{3}}{2} + \left(-\frac{1}{2}\right)\cdot\left(-\frac{\sqrt{3}}{2}\right) + \frac{1}{2}\cdot\frac{\sqrt{3}}{2} + \frac{1}{2}\cdot\left(-\frac{\sqrt{3}}{2}\right) - 1\cdot 0 = 0 = \text{RHS} = \frac{[g]}{n_\mu}$$

$$= \frac{6}{2}\delta^{11}\delta_{11}\delta_{12}.$$

We'd get similar results using a three-dimensional 3×3 matrix representation for D_3:

$$e = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad c = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad c^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$b_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad b_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Let, say, $i = j = r = s = 1$. Then

$$\sum_g D_{ir}^{(\mu)}(g) D_{js}^{(\nu)}(g) = D_{11}^{(\mu)}(e) D_{11}^{(\mu)}(e^{-1}) + D_{11}^{(\mu)}(c) D_{11}^{(\mu)}(c^{-1}) + D_{11}^{(\mu)}(c^2) D_{11}^{(\mu)}\left(c^{2^{-1}}\right)$$

$$+ D_{11}^{(\mu)}(b) D_{11}^{(\mu)}(b^{-1}) + D_{11}^{(\mu)}(b_2) D_{11}^{(\mu)}(b_2^{-1}) + D_{11}^{(\mu)}(b_3) D_{11}^{(\mu)} b_3^{-1})$$

$$= 1 + 0 + 0 + 0 + 0 + 1 = 2 = \text{RHS} = \frac{[g]}{n_\mu} = \frac{6}{3} \delta^{11} \delta_{11} \delta_{11}.$$

What about $i = j = r = 1;\ s = 2$? Then

$$\sum_g D_{ir}^{(\mu)}(g) D_{js}^{(\nu)}(g) = D_{11}^{(\mu)}(e) D_{12}^{(\mu)}(e^{-1}) + D_{11}^{(\mu)}(c) D_{12}^{(\mu)}(c^{-1}) + D_{11}^{(\mu)}(c^2) D_{12}^{(\mu)}\left(c^{2^{-1}}\right)$$

$$+ D_{11}^{(\mu)}(b) D_{12}^{(\mu)}(b^{-1}) + D_{11}^{(\mu)}(b_2) D_{12}^{(\mu)}(b_2^{-1}) + D_{11}^{(\mu)}(b_3) D_{12}^{(\mu)} b_3^{-1})$$

$$= 0 + 0 + 0 + 0 + 0 + 0 = 0 = \text{RHS} = \frac{[g]}{n_\mu} = \frac{6}{3} \delta^{11} \delta_{11} \delta_{12}.$$

What about $i = j = 1;\ r = s = 2$? Then

$$\sum_g D_{ir}^{(\mu)}(g) D_{js}^{(\nu)}(g) = D_{11}^{(\mu)}(e) D_{22}^{(\mu)}(e^{-1}) + D_{11}^{(\mu)}(c) D_{22}^{(\mu)}(c^{-1}) + D_{11}^{(\mu)}(c^2) D_{22}^{(\mu)}\left(c^{2^{-1}}\right)$$

$$+ D_{11}^{(\mu)}(b) D_{22}^{(\mu)}(b^{-1}) + D_{11}^{(\mu)}(b_2) D_{22}^{(\mu)}(b_2^{-1}) + D_{11}^{(\mu)}(b_3) D_{22}^{(\mu)} b_3^{-1})$$

$$= 1 + 0 + 0 + 0 + 0 + 1 = 2 = \text{RHS} = \frac{[g]}{n_\mu} = \frac{6}{3} \delta^{11} \delta_{11} \delta_{22}.$$

Without explicitly plugging in, I get a headache with the abstract linear algebra.

Restriction on the number of irreducible representations. With what we have just learned there is no loss in generality if we let the representations of $D^{(\mu)}$ and $D^{(\nu)}$ be unitary (because we can always make a unitary scalar product.) Then our result above becomes

$$\sum_g D_{ir}^{(\mu)}(g) D_{js}^{(\nu)^*}(g) = \frac{[g]}{n_\mu} \delta^{\mu\nu}\, \delta_{ij} \delta_{rs}.$$

Don't let your eyes not see the star up there thanks to Maschke's theorem.

Fix your attention on $D^{(\mu)}$ and set $\nu = \mu$. For fixed i and r the set of objects $\{D_{ir}^{(\mu)}(g_1), \ldots, D_{ir}^{(\mu)}(g_{[g]})\}$ can be considered a $[g]$-dimensional column vector. Then the LHS of the sum represents the complex scalar product of two vectors in this space, labeled by the pairs of indices (ir) and (js) respectively. Each of these indices takes values from 1 to n_μ, the dimensionality of $D^{(\mu)}$. There are n_μ^2 of these vectors, and by the sum (scalar product) above, they are all orthogonal. The same applies for any other value of μ, say μ', and moreover, the vectors formed by the $D^{(\mu')}$ will be orthogonal to those formed by the $D^{(\mu)}$. Accounting for all possible values of μ we can form a total of $\sum_\mu n_\mu^2$ mutually orthogonal vectors. The number of such vectors cannot exceed the dimensionality of the space, this being $[g]$. We have therefore proved that $\sum_\mu n_\mu^2 \leq [g]$. (Later this will be shown to be an equality.) Since each n_μ must be at least 1, the number of irreducible representations of a finite group is strictly limited. Want to understand this? Plug in our D_3 example.

Orthogonality of characters. Recall that the character of a representation D is the set $\{\chi(g)\}$ where $\chi(g)$ is the trace of the matrix $D(g)$. From linear algebra, the traces also have the properties:

χ is the same for equivalent representations connected by a similarity $D'(g) = SD(g)S^{-1}$.

1-χ is the same for conjugate elements (of the same conjugacy class) since

2-$D(hgh^{-1}) = D(h)D(g)\big(D(h)\big)^{-1}$.

3-If D is unitary ($D^{-1} = D$^ †) then$\chi(g^{-1}) = \mathrm{Tr}\left(\big(D(g)\big)^{-1}\right) = \mathrm{Tr}(D(g') †) = \chi^*(g)$. (This result is always true for finite or compact groups since any representation is equivalent to a unitary representation.)

The orthogonality relation for characters is obtained by taking suitable traces of the fundamental orthogonality theorem—the orthogonality results for the matrices pass through to their traces, and further constrain our algebra. Tracing over i, r and s, j (multiplying by $\delta_{ir}\,\delta_{sj}$ gives

$$\sum_g \chi^{(\mu)}(g)\chi^{(\nu)}(g^{-1}) = \frac{[g]}{n_\mu}\delta^{\mu\nu}\delta_{ij}\delta_{ij}.$$

Here $\delta_{ij}\delta_{ij} = \delta_{ii} = n_\mu$. The summation convention is always operative as the indices range from 1 to n_μ, the dimension of the irreducible representation. Thus,

$$\frac{1}{[g]}\sum_g \chi^{(\mu)}(g)\chi^{(\nu)}(g^{-1}) = \delta^{\mu\nu}.$$

By virtue of the third property we listed for the traces of χ, this sum can be recast in the alternative form

$$\frac{1}{[g]}\sum_g \chi^{(\mu)}(g)\chi^{(\nu)^*}(g) = \delta^{\mu\nu}.$$

Thus up to a factor of $[g]$, the LHS of the equation is the usual complex scalar product of the two $[g]$-dimensional column vectors $\left(\chi^\mu(g_1), \chi^\mu(g_2), \dots, \chi^\mu\left(g_{[g]}\right)\right)$. So it's both convenient and illuminating to define the scalar product of two characters φ, χ by

$$\langle\varphi, \chi\rangle = \frac{1}{[g]}\sum_g \phi(g)\chi(g^{-1}) = \langle\chi, \varphi\rangle.$$

In this language $\frac{1}{[g]}\sum_g \chi^{(\mu)}(g)\chi^{(\nu)}(g^{-1}) = \delta^{\mu\nu}$ states that the characters of inequivalent irreducible representations are orthonormal: $\langle\chi^{(\mu)}, \chi^{(\nu)}\rangle = \delta^{\mu\nu}$.

Given that the characters of conjugate elements are equal, all elements of a conjugacy class have the same character, so the distinct characters may be labeled as $\chi_i, i = 1, \dots, k$. These correspond to ht k

conjugacy classes K_i . Let k_i be the number of elements in the conjugacy class K_i. Then the sum over g in $\frac{1}{[g]}\sum_g \chi^{(\mu)}(g)\chi^{(\nu)^*}(g) = \delta^{\mu\nu}$ can be rewritten as a sum over i:

$$\frac{1}{[g]}\sum_i k_i \chi_i^{(\mu)} \chi_i^{(\nu)} = \delta^{(\mu\nu)}.$$

But this can now be interpreted as orthogonality of the vectors $\sqrt{k_i}\chi_i^{(\mu)}$ (no summation) in k-dimensional space. And since there are no more than k such vectors, we have another inequality on the number r of different irreducible representations: $r \leq k$. (This turns into an equality (See "Group Theory and its Applications to Physical Problems," M Hamermesh, Addison-Wesley, out of print.)) The characters can be shown to be orthogonal wrt the index i as well as in the sense that

$$\frac{1}{[g]}\sum_\mu k_i \chi_i^{(\mu)} \chi_j^{(\nu)^*} = \delta_{ij}.$$

This gives the inequality in the opposite direction, thus leading us to the inescapable conclusion that $r = k$.

Decomposition of Reducible Representation. For a finite or compact group, any reducible representation is reducible into the direct sum of irreducible representations, meaning that the representation matrices can be put in block diagonal form, the nonzero diagonal blocks being the matrices of the irreducible representations. A given representation may appear more than once, *e.g.*, a 5-dimensional representation could decompose into a trivial representation and two copies of the same 2-dimensional representation. Generally we may write

$$D = \sum_{\oplus\nu} a_\nu D^{(\nu)},$$

where the non-negative integers a_ν denote the number of times a particular irreducible representation of $D^{(\nu)}$ appears in the decomposition. Let's go find these a_ν. We will do so by using the orthogonality properties of the characters $\chi^{(\nu)}$. Taking the trace of both sides of the sum above for an arbitrary group element g, we can see that the characters $\chi(g)$ of D decomposes as an ordinary sum of the characters $\chi^{(\nu)}$:

$$\chi(g) = \sum_g a_\nu D^{(\nu)}.$$

Let's dub χ a compound character, and the $\chi^{(\nu)}$ simple characters. Multiply the sum above by $\chi^{(\mu)}(g^{-1})$ and sum over g to get

$$\sum_g \chi^{(\mu)}(g^{-1})\chi(g) = \sum_\nu a_\nu \sum_g \chi^{\mu}(g^{-1})\chi^{(\nu)}(g).$$

Using $\frac{1}{[g]}\sum_g \chi^{(\mu)}(g)\chi^{(\nu)}(g^{-1}) = \delta^{\mu\nu}$, the RHS is just $\sum_\nu a_\nu [g]\delta^{\mu\nu} = [g]a_\mu$. Thus

$$a_\mu = \frac{1}{[g]}\sum_g \chi(g)\chi^{(\mu)}(g^{-1}).$$

All of this amounts to writing

$$\chi = \sum_g a_\nu \chi^{(\nu)}$$

and taking the scalar product $\langle \chi^{(\mu)}, \chi \rangle$ to obtain

$$a_\mu = \langle \chi\text{^}((\mu), \chi \rangle$$

which is the shorthand version of the sum for a_μ.

Regular representation. Recall Cayley's theorem? There is an isomorphism between a group G and a subgroup of the permutation group $S_{[g]}$ provided by left multiplication. That is

$$g g_i = \sum_j D_{ji}(g) g_j.$$

The $[g] \times [g]$ permutation matrices $D_{ij}(g)$form a $[g]$-dimensional representation of a group called ***regular representation***. Now $g g_i$ is an element of G, say it's $g_{l,}$ $l \neq i$ except when $g = e$. We see that $D_{ji}(g)$ has only one non-zero element in each row and each column. For $g \neq e$, all these elements are off diagonal ($l \neq i$), while for $g = e$, they are all diagonal. In fact $D_{ji}(e)$is the unit matrix δ_{ji}.

Example 6—In C_3

C_3	$e = g_1$	$c = g_2$	$c^2 = g_3$
$D^{(1)}$	1	1	1
$D^{(2)}$	1	ω	ω^2
$D^{(3)}$	1	ω^2	ω

Recall that

$$D(c) = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \qquad D(e) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; D(c^2) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

This is the regular representation for C_3, there the dimensions of the matrix match the order (size) of C_3.

Now we decompose the regular representation into its irreducible components $D = \sum_{\oplus} a_\nu D^{(\nu)}$. The coefficients a_μ are given by $a_\mu = \langle \chi^{(\mu)}, \chi \rangle$:

$$a_\mu = \frac{1}{[g]} \sum_g \chi(g) x^{(\mu)}(g^{-1}).$$

From the special form of the D_μ we deduce

$$\chi(g) = \begin{cases} 0, & g \neq e \\ [g], & g = e. \end{cases}$$

Therefore $a_\mu = \chi^{(\mu)}(e) = n_\mu$, the dimensionality of $D^{(\mu)}$. Putting $g = e$ in

$$\chi(g) = \sum_\nu a_\nu \chi^{(\nu)}(g)$$

results in

$$[g] = \sum_\nu n_\nu \cdot n_\nu = \sum_\nu n_\nu^2$$

making $\sum_\mu n_\mu^2 \leq g$ an equality. For $g \neq e$, we instead get

$$0 = \sum_\nu n_\nu \chi^{(\nu)}(g)$$

and the last two equations can be combined as

$$\frac{1}{[g]} \sum_\nu \chi^{(\nu)}(e) \chi^{(\nu)}(g) = \begin{cases} 0, & g \neq e \\ 1, & g = e. \end{cases}$$

So what we did in "The example (5)" was to use the tools:

1. The number of irreducible representations equals the number of conjugacy classes.
2. $\sum_\mu n_\nu^2 = [g]$.
3. Orthogonality: $\sum k_i \chi_i^{(\mu)} \chi_i^{(\nu)^*} = [g]$.
4. Whatever tricks we can use.

Example 7—(Please compare this example with the equivalent material in Georgi's text.) The character table of D_3. The conjugacy classes for this group were $k_1 = [e]$, $k_2 = [c, c^2]$, and $k_3 = [b, bc, bc^2 = b, b_2, b_3]$. Tool 2 states that $n_1^2 + n_2^2 + n_3^2 = 6$, but there is always a trivial group of size 1. This leaves $n_2^2 + n_3^2 = 5$. Eyeballs tell you, WLOG, that $n_2 = 1$ and $n_3 = 2$. This enables us to fill the character table in since $\chi^{(\mu)}(e) = n_\mu$. For the one-dimensional representation the $\chi's$ must mimic the group structure, so $\chi(bc) = \chi(b)\chi(c)$. Note that $\chi(b) = \chi(bc) = \chi_3$. Thus $\chi(c) = \chi_2 = 1$. Further, $\chi(b)^2 = \chi(b^2) = \chi(e) = 1$, giving $\chi_3 = -1$ for $D^{(2)}$. Using (1), (2) and (4) gets us this:

D_3	k_1	k_2	k_3
$D^{(1)}$	1	1	1
$D^{(2)}$	1	1	-1
$D^{(3)}$	2	α	β

Let's use orthogonality (3). The orthogonality of $\chi^{(3)}$ with $\chi^{(1)}$ and $\chi^{(2)}$ gives(via dot product):

$$2 + 2\alpha + 3\beta = 0$$

$$2 + 2\alpha - 3\beta = 0.$$

The coefficients in the equations are the multiplicities of k_1 and k_2 of the respective conjugacy classes. The equations yield $\alpha = -1$ and $\beta = 0$. (The three characters should be orthonormal.)

We need the matrix of at least one of the bs. WLOG let b be rotation about the x-axis:

$$D^{V=\mathbb{R}^3} = \mathrm{diag}(1, -1, -1).$$

This corresponds to $x' = x,\ \ y' = y,\ \ z' = -z$. $\mathrm{Tr}D^V(b) = \chi^V = -1$. So $\chi^V = (3,0,-1)$. The 3 is the trace of the identity element in conjugacy class $K^{(1)}$, the 0 is the trace of the conjugacy class $[c.c^2]$, and the -1 is the trace of the conjugacy class of $[b, b_2, b_3]$. Thus

$$\chi^{V=\mathbb{R}^3} = \chi^{(2)} + \chi^{(3)}, \qquad \text{hence,} \qquad D^V = D^{(2)} \oplus D^{(3)}.$$

The basis functions of $D^{(2)}$ and $D^{(3)}$ can be reasoned out. We've noted that z changes sign under b, but it's left invariant under c. Hence z forms the basis for $D^{(2)}$. The two-dimensional representation $D^{(3)}$ gives the transformation of x and y, but unlike for C_3, and precisely because of the rotations b, b_2 and b_3, these transformations can't be further broken reduced. Let's do some checking:

$$a_\mu = \frac{1}{[g]}\sum_g \chi(g)\chi^{(\mu)}(g^{-1}) = \sum_g k_i\chi^{(\mu)}(g^{-1})\chi(g):$$

$$a_1 = \frac{1}{6}\begin{bmatrix}1\\2\\3\end{bmatrix}[1\ 1\ 1]\begin{bmatrix}3\\0\\-1\end{bmatrix} = \frac{1}{6}[1(1)+2(1)+3(1)]\begin{bmatrix}3\\0\\-1\end{bmatrix} = 0, \qquad D^{(1)}\text{row in table; } 1-\text{dimensional.}$$

$$a_2 = \frac{1}{6}[1(1)+2(1)+3(-1)]\begin{bmatrix}3\\0\\-1\end{bmatrix} = 1, \qquad D^{(2)}\text{row in table.; } 2-\text{dimensional.}$$

$$a_3 = \frac{1}{6}[1(1)+2(1)+3(0)]\begin{bmatrix}3\\0\\-1\end{bmatrix} = 1, \qquad D^{(3)}\text{row in table, } 3-\text{dimensional.}$$

So $D^{V=\mathbb{R}^3} = D^{(2)} \oplus D^{(3)} = \sum_\oplus a_\nu D^{(\nu)}$. $\chi = \sum_\nu a_\nu \chi^{(\nu)} = 0\cdot\chi^{(1)} + 1\cdot\chi^{(2)} + 1\cdot\chi^{(3)}$.

Please see the Jones text, chapter 5 for applications to Ferromagnetism and Ferroelectricity ($D^V = D^{(1)} \oplus D^{(2)} \oplus D^{(3)}$.) Also in this chapter is a study of the molecular vibrations of water, and the breaking of degeneracy by the addition of an interaction Hamiltonian to the "free" Hamiltonian which is invariant under some group G. If you have a background in quantum physics in Dirac's bra-ket formalism (there is a quick review in the appendix of Jones) you have sufficient background to understand the first chapter of Georgi's text, which contains the identical material BUT in Dirac's bra-ket notation, and the definition of projection operators together with a great physical example of the power of these methods for discrete groups. The remainder of Georgi then tackles continuous (Lie) groups, this material being paralleled with Jones from chapter 6 to its end. This latter material is what you need to grasp not only the Standard Model of physics, but its generalizations starting with SU(5) grand unification theory (the SU(5) GUT). I strongly recommend you read the group theoretic material (in the appendix as well) of Michio Kaku's text, "Quantum Field Theory, A Modern Introduction," Oxford, 1993. Review and make sure you understand the quantum mechanics of spin and angular momenta, spinors and tensors at the level of "Quantum Mechanics," Vols. I and II by C. Cohen-Tannoudji, B. Diu, and F. Laloë, A Wiley-Interscience Publication, 1977.

Thankfully, only two more algebraic blocks remain in these notes. The first chapter of Gilmore's 2008 text tying what we've just studied to Galois theory follows next. This is proceeded by a few notes I wrote to make "Lie Algebras in Particle Physics," by H. Georgi more readable starting with his chapter 2. I consolidated his sloppy book with the appropriate chapters from the treatment of continuous groups in Jones (chapter 6-end), and may one day write up these latter notes. For now we turn to Galois theory. The coverage of this material will be fast paced. Then it's on to the last notes on group theory covering the adjoint representation of the Gell-Mann matrices.

Galois theory.

Galois' theorem: A polynomial over the complex field can be solved iff its Galois group G contains a chain (tower) of subgroups $G = G_0 \supset G_1 \supset \cdots \supset G_\omega = \mathbb{I} \ni G_{i+1} \trianglelefteq G_i$ and G_i/G_{i+1} is abelian.

Quadratic Equation. Take $(z - r_1)(z - r_2) = 0$. This is $z^2 - I_1 z + I_2$, where $I_1 = r_1 + r_2$, and $I_2 - r_1 r_2$. Both **invariants** I_1, I_2 are symmetric under the permutation group S_2 with character table:

S_2	$\mathbb{I}$	(12)
Γ^1	1	1
Γ^2	1	-1

We use the character table to make linear combinations of r_1 and r_2:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} r_1 + r_2 \\ r_1 - r_2 \end{pmatrix}$$

The sum is invariant under the identity $\mathbb{I}$, but the difference $r_1 - r_2$ is not invariant under the two-cycle $(1\ \ 2)$: $(r_1 - r_2) = -(r_1 - r_2)$. ***But*** S_2 is abelian (it "increases the symmetry" if you will). There exist a function of two symbols symmetric under $(1\ \ 2)$ because of the commutativity of S_2. Let's try $(r_1 - r_2)^2$.

$$(r_1 - r_2)^2 = r_1^2 - 2r_1 r_2 + r_2^2 = r_1^2 + 2r_1 r_2 + r_2^2 - 4r_1 r_2 = I_i^2 - 4I_2 = D,$$

where D is the discriminant of the quadratic equation. Now check out the action of the character table rephrased:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} I_1 \\ \pm\sqrt{I_1 - 4I_2} \end{pmatrix}.$$

Note that the RHS column vector is symmetric under both the identity $\mathbb{I}$ and S_2, moreover both solutions can now be seen as functions of the two invariants I_1 and I_2. We have found solutions by adding/subtracting, multiplying/dividing and extracting roots. This is what is meant for a polynomial to be solvable by "radical extension."

$$r_1, r_2 = \frac{1}{2}\left(I_1 \pm \sqrt{D}\right).$$

The character table, by the way, is an orthonormal rotation matrix which preserves lengths. The Galois group is the abelian group S_2. It has a ***tower*** (or ***chain***) ending in the identity element:

$$\begin{array}{c} S_2 \\ \downarrow \\ e \end{array}$$

If you recall $e \,\underline{\Delta}\, S_2$. We will see that the tower of the group of a given polynomial will have much to do with its solvability.

<u>Cubic Equation</u>. Take $(z - s_1)(z - s_2)(z - s_3) = z^3 + I_1 z^2 + I_3 z - I_3 = 0$, where

$$I_1 = s_1 + s_2 + s_3, \qquad I_2 = s_1 s_2 + s_1 s_3 + s_2 s_3, \qquad I_3 = s_1 s_2 s_3.$$

The Galois group is S_3 with the subgroup chain as shown:

$$\begin{array}{ccccc} & & S_3 = D_3 & & \\ & \swarrow & & \searrow & \\ A_3 = C_3 & & \downarrow & & S_2 \\ & \searrow & & \swarrow & \\ & & e & & \end{array}$$

Note that $e \,\underline{\Delta}\, A_3 \,\underline{\Delta}\, S_3$. The existence of this tower is the first of two conditions of Galois' theorem on the solvability of polynomials by radical extension. I'm sneaking in the theorem by example. Also note that $S_3/A_3 = S_2$ is abelian, as is $A_3/e = A_3$. This is the second condition of Galois' famous theorem. Now we know that there is a solution to the cubic by radical extension.

We begin with the solution of the last group-subgroup pair in the chain (or tower). We know how to derive the character table for $C_3 = A_3$ (recall "The BIG example (5)":

$A_3 = C_3$	$e = \mathbb{I}$	$c = (123)$	$c^2 = (321)$
$\Gamma^{(1)}$	1	1	1
$\Gamma^{(2)}$	1	ω	ω^2
$\Gamma^{(3)}$	1	ω^2	ω

Since $\omega^3 = 1$, recall that there were three roots of unity. Note also that The Γ^i form a group isomorphic with $A_3 = C_3$. Linear combinations of the roots that transform under the three one-dimensional irreducible representations are easily constructed using the character table:

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} s_1 + s_2 + s_3 \\ s_1 + \omega s_2 + \omega^2 s_3 \\ s_1 + \omega^2 s_2 + \omega s_3 \end{bmatrix}.$$

For example, the action of $(1\ \ 2\ \ 3)^{-1}$ on v_2 is

$$(1\ \ 2\ \ 3)^{-1}(s_1 + \omega s_2 + \omega^2 s_3) = (3\ \ 2\ \ 1)(s_1 + \omega s_2 + \omega^2 s_3) = \omega(s_1 + \omega s_2 + \omega^2 s_3) = \omega v_2.$$

Since v_1 is symmetric under both A_3 and S_3 it can be expressed in terms of the invariants I_k: $v_1 = I_1$. The remaining functions v_2 and v_3 are symmetric under $\mathbb{I}$ but not umnder A_3.

Proceed to the next group-subgroup pair $S_3 \supset A_3$ as before. To construct functions symmetric under A_3 but not under S_3 note that the cubes of v_2 and v_3 are symmetric under A_3 but not under S_3:

$$(1\ \ 2)v_2^3 = (1\,2)(s_1 + \omega s_2 + \omega^2 s_3)^3 = (s_2 + \omega s_1 + \omega^2 s_3)^3 = \omega^3(s_1 + \omega^2 s_2 + \omega s_3)^3 = v_3^3,$$

$$(1\,2)v_3^3 = (1\,2)(s_1 + \omega^2 s_2 + \omega s_3)^3 = \left(s_2 + \omega^2 s_{s_1} + \omega s_3\right)^3 = \omega^6(s_1 + \omega s_2 + \omega^2 s_3) = v_2^3.$$

Since $S_2 = S_3/A_3$ permutes the functions v_2^3 and v_3^3, it is the Galois group of the resolvent quadratic equation whose two roots are v_2^3 and v_3^3. This equation has the form:

$$(x - v_2^3)(x - v_3^3) = x^2 - J_1 x + J_2 = 0,$$

$$J_1 = v_2^3 + v_3^3, \qquad J_2 = v_2^3 v_3^3.$$

(We know how to solve quadratic equations.) Take note that J_1 and J_2 are symmetric under S_3. They can then be expressed in terms of the invariants I_1, I_2 and I_3 of the original cubic equation as follows:

$$J_1 = \sum_{i+2j+3k=3} A_{ijk} I_1^i I_2^j I_3^k,$$

$$J_2 = \sum_{1+2j+3k=6} B_{ijk} I_1^i I_2^j I_3^k.$$

This sum can be greatly simplified by letting $I_1 = s_1 + s_2 + s_3 = 0$. To do this, shift the origin with $z = y + \frac{1}{3} I_1$. Then the auxiliary cubic equation has structure

$$y^3 - 0y^2 + I_x' y - I_3' = 0 \text{ with}$$

$$I_1' = s_1' + s_2' + s_3' = 0,$$

$$I_2' = s_1' s_2' + s_1' s_3' + s_2' s_3' = I_2 - \frac{1}{3} I_1^2,$$

$$I_3' = s_1' s_2' s_3' = I_3 - \frac{1}{3} I_2 I_1 + \frac{2}{27} I_1^3.$$

The two invariants $J_1 = v_2^3 + v_3^3$ and $J_2 = v_2^3 v_3^3$ can then be expressed in terms of the invariants I_2' and I_3' as follows:

$$J_1 = v_2^3 + v_3^3 - -27 I_3', \qquad J_2 = v_2^3 v_3^3 = -27 I_2'^3.$$

The resolvent quadratic equation whose solution provides v_2^3, v_3^3 is $x^2-(-27I_3')x+(-27I_x'^3)=0$.

The two solutions to this resolvent equation are$v_2^3, v_3^3 = -\frac{27}{2}I_3' \pm \frac{1}{2}[(27I_3')^2+4\times 27I_2'^3]^{\frac{1}{2}}$. The roots v_2 and v_3 are obtained by taking the cube roots of v_2^3 and v_3^3:

$$v_2, v_3 = \left\{-\frac{27}{2}I_3' \pm \frac{1}{2}[(27I_3')^2+4\times 27I_2'^3]^{\frac{1}{2}}\right\}^{\frac{1}{3}}.$$

Finally the roots s_1, s_2 and s_3 are linearly related to v_1, v_2 andv_3 by:

$$\begin{bmatrix}1 & 1 & 1\\ 1 & \omega & \omega^2\\ 1 & \omega^2 & \omega\end{bmatrix}\begin{bmatrix}s_1\\ s_2\\ s_3\end{bmatrix}=\begin{bmatrix}v_1\\ v_2\\ v_3\end{bmatrix}=\begin{bmatrix}v_1+v_2+v_3\\ v_1+\omega^2 v_2+\omega v_3\\ v_1+\omega v_2+\omega^2 v_3\end{bmatrix}.$$

Quartic equation. The general quartic has form

$$(z-t_1)(z-t_2)(z-t_3)(z-t_4)=z^4-I_1z^3+I_2z^2-I_3z+I_4=0,$$

$$I_1=t_1+t_2+t_3+t_4, \qquad I_2=t_1t_2+t_1t_3+t_1t_4+t_2t_3+t_2t_4+t_3t_4,$$

$$I_3=t_1t_2t_3+t_1t_2t_4+t_1t_3t_4+t_2t_3t_4, \qquad I_4=t_1t_2t_3t_4.$$

For later convenience we will construct the auxiliary quartic by shifting the origin of coordinates by $z=z'+\frac{1}{4}I_1$ (Tschirnhaus transformation). The

$$(z'-t_1)(z'-t_2)(z'-t_3)(z'-t_4)=z'^4-I_1'z'^3+I_2'z'^2-I_3'z'+I_4'=0,$$

$$I_1'=0,$$

$$I_2'=I_2-\frac{3}{8}I_1^2$$

$$I_3'=I_3-\frac{1}{2}I_2I_1+\frac{1}{8}I_3^3$$

$$I_4' = I_4 - \frac{1}{4}I_3I_1 + \frac{1}{16}I_2I_1^2 - \frac{3}{4^4}I_1^4.$$

The Galois group is S_4. It looks like:

$$\begin{array}{ccccc} & & S_4 & & \\ & \swarrow & \downarrow & \searrow & \\ V_8 & & A_4 & & S_3 \\ \downarrow & \swarrow & & & \downarrow \\ V_4 & & \downarrow & & S_2 \\ & \searrow & & \swarrow & \\ & & \mathbb{I} & & \end{array}$$

The tower (or chain) is $e \underline{\Delta} V_4 \underline{\Delta} A_4 \underline{\Delta} S_4$. ($V_4$ is the vieregruppe, Klein group, or Klein four-group: { $\mathbb{I}$,(1 2)(3 4),(1 3)(2 4),(1 4)(2 3)}. This satisfies both conditions of Galois' theorem:

(i) A_4 is invariant (normal) in S_4 and $\frac{S_4}{A_4} = S_2$.

(ii) V_4 is invariant (normal) in A_4 and $\frac{A_4}{V_4} = C_3 = \{\mathbb{I}, (2\ 3\ 4), (4\ 3\ 2)\}$.

(iii) $\mathbb{I}$ is invariant in V_4 and $\frac{V_4}{\mathbb{I}} = V_4 = \{\mathbb{I}, (1\ 2)(3\ 4), (1\ 3)(2\ 4), (1\ 4)(2\ 3)\}$.

As before, begin at the end of the chain with the abelian group V_4 with character table given by:

	$\mathbb{I}$	$(1\ 2)(3\ 4)$	$(1\ 3)(2\ 4)$	$(1\ 4)(2\ 3)$	Basis functions
Γ^1	1	1	1	1	$\omega_1 = t_1 + t_2 + t_3 + t_4$
Γ^2	1	1	-1	-1	$\omega_2 = t_1 + t_2 - t_3 - t_4$
Γ^3	1	-1	1	-1	$\omega_3 = t_1 + t_2 - t_3 - t_4$
Γ^4	1	-1	-1	1	$\omega_4 = t_1 - t_2 - t_3 + t_4$

The linear combination of these roots that transform under each of the irreducible representations are

$$\begin{bmatrix}\omega_1\\ \omega_2\\ \omega_3\\ \omega_4\end{bmatrix} = \begin{bmatrix}1 & 1 & 1 & 1\\ 1 & 1 & -1 & -1\\ 1 & -1 & 1 & -1\\ 1 & -1 & -1 & 1\end{bmatrix}\begin{bmatrix}t_1\\ t_2\\ t_3\\ t_4\end{bmatrix} = \begin{bmatrix}t_1+t_2+t_3+t_4\\ t_1+t_2-t_3-t_4\\ t_1+t_2-t_3-t_4\\ t_1-t_2-t_3+t_4\end{bmatrix}.$$

These basis vectors are symmetric under $\mathbb{I}$ but the basis vectors ω_2, ω_3 and ω_4 are not symmetric under V_4. Thus we now move up the chain to $A_4 \supset V_4$.

We seek to construct from the above linear combinations functions that are

(i) Symmetric under V_4

(ii) Permuted among themselves by A_4 and the group A_4/V_4.

These functions are $\omega_1 = \mathbb{I}$ and $\omega_2^2, \omega_3^2, \omega_4^2$. In the coordinate system in which the sum of the roots is zero, the three functions $\omega_2^2, \omega_3^2, \omega_4^2$ are

$$\omega_2^2 = (t_1' + t_2' + t_3' + t_4')^2 = 2^2(t_1' + t_2')^2 = -4(t_1' + t_2')(t_3' + t_4'),$$

$$\omega_3^2 = (t_1' - t_2' + t_3' - t_4')^2 = 2^2(t_1' + t_3') = -4(t_1' + t_3')(t_2' + t_4'),$$

$$\omega_4^2 = (t_1' - t_2' - t_3' + t_4')^2 2^2(t_1' + t_4')^2 = -4(t_1' + t_4')(t_2' + t_3').$$

It's clear that the $\omega_2^2, \omega_3^2, \omega_4^2$ are permuted among themselves by the factor group $C_3 = A_4/V_4$, which is a subgroup of the Galois group of a resolvent cubic equation whose three roots are $\omega_2^2, \omega_3^2, \omega_4^2$:

$$(y - \omega_2^2)(y - \omega_3^2)(y - \omega_4^2) = y^3 - J_1 y^2 + J_2 y - J_3 = 0,$$

$$J_1 = \omega_2^2 + \omega_3^2 + \omega_4^2,\ \ J_2 = \omega_2^2\omega_3^2 + \omega_2^2\omega_4^2 + \omega_3^2\omega_4^2,\ \ J_3 = \omega_2^2\omega_3^2\omega_4^2.$$

Since the J's are invariant under C_3, they can be expressed in terms of the symmetric functions (coefficients) of the original quartic (either one):

$$J_1 = (-4)^1(2I_2'),$$

$$J_2 = (-4)^2(I_2'^2 - 4I_4'^2),$$

$$J_3 = (-4)^3(-I_3'^3).$$

This cubic equation is solved by proceeding to the first group-subgroup pair in the chain: $S_4 \supset A_4$ with $\frac{S_4}{A_4} = S_2$. The cubic is solved by introducing the resolvent quadratic of the previous section. If the three solutions of the resolvent cube are called y_1, y_2, y_3, then the functions $\omega_2^2, \omega_3^2, \omega_4^2$ are

$$\omega_2 = \pm\sqrt{y_2}, \qquad \omega_3 = \pm\sqrt{y_3}, \qquad \omega_4 = \pm\sqrt{y_4}.$$

A simple calculation shows that $\omega_2\omega_3\omega_4 = 8I_3'$. The signs $\pm\sqrt{y_i}$ are selected so that their product is $8I_3'$. The simple linear relation between the roots t_i and the invariants I_1 and functions $\omega_j(I')$ is easily inverted:

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} I_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \end{bmatrix}$$

where the ω_j are square roots of the solutions of the resolvent cubic equation whose coefficients are functions of the auxiliary quartic.

Insolvability of the quintic. The Galois group for the general quintic is S_5. It's tower is $\mathbb{I} \trianglelefteq A_5 \trianglelefteq S_5$. Although S_5/A_5 is commutative, here is the rub: $A_5/\mathbb{I}$ is not abelian. This tower doesn't satisfy Galois' theorem. Let's discuss this. Given a polynomial of n^{th} degree in a factorized form, it is trivial to construct invariants that are symmetric with S_n, *e.g.*, for $n = 5$,

$$(z - r_1)(z - r_2)(z - r_3)(z - r_4)(z - r_5) = 0, \qquad I_1 = r_1 + r_2 + r_3 + r_4 + r_5,$$

$$I_2 = r_1r_2 + r_1r_3 + r_1r_4 + r_1r_5 + r_2r_3 + r_2r_4 + r_2r_5 + r_3r_4 + r_3r_5 + r_4r_5,$$

$$I_3 = r_1r_2r_3 + r_1r_2r_4 + r_1r_2r_5 + r_2r_3r_4 + r_2r_3r_5 + r_3r_4r_5,$$

$$I_4 = r_1r_2r_3r_4 + r_1r_2r_3r_5 + r_2r_3r_4r_5, \qquad I_5 = r_1r_2r_3r_4r_5.$$

However, we often only know the coefficientsI_i, $i = 1, \dots, n$, *e.g.*, $y^5 - 3y^4 + 2y^3 - y^2 - \frac{1}{2}y = 0$. The task, then, is to find the n roots by radical extension as we have done for the quadratic, cubic and quartic given only the coefficients I_i. This is a constructive process called ***field extension***, in which we seek solutions by using arithmetic operators and raising quantities to positive integer powers. In this sense, as with the restriction of geometric constructions to straightedge and compass, we quickly run into things we can't construct (polynomials we can't solve by radicals).

The problem with the quintic is that we know (if given five roots r_1, r_2, r_3, r_4, r_5) that we may construct five invariants (see above) symmetric wrt S_5. When we know only the coefficients I_i, the process to find the five roots begins at the group-subgroup pair ending with the identity: $\mathbb{I} \trianglelefteq A_5$ to construct invariant functions that are invariant with $\mathbb{I}$ but not with A_5. The character table for A_5 produces a set of basis functions that are invariant under $\mathbb{I}$ but not with A_5. Then we seek to construct from these basis functions linear combinations that are symmetric under A_5 and permuted among by the group S_5/A_5. But A_5 is nonabelian, rendering it impossible to construct invariants symmetric with A_5 using addition/subtraction, multiplication/division, raising by integer powers. Let's make this crystal clear. Back at the cubic the character table was:

$A_3 = C_3$	$e = \mathbb{I}$	$c = (123)$	$c^2 = (321)$
$\Gamma^{(1)}$	1	1	1
$\Gamma^{(2)}$	1	ω	ω^2
$\Gamma^{(3)}$	1	ω^2	ω

Did you notice the Γ_i's form a group?

$\cong A_3$	$\Gamma^{(1)}$	$\Gamma^{(2)}$	$\Gamma^{(3)}$
$\Gamma^{(1)}$	$\Gamma^{(1)}$	$\Gamma^{(2)}$	$\Gamma^{(3)}$
$\Gamma^{(2)}$	$\Gamma^{(2)}$	$\Gamma^{(3)}$	$\Gamma^{(1)}$
$\Gamma^{(3)}$	$\Gamma^{(3)}$	$\Gamma^{(1)}$	$\Gamma^{(2)}$

Under a scalar product, $\Gamma^{(2)}\Gamma^{(3)} = (1+\omega+\omega^2)\cdot(1+\omega^2+\omega) = 1+\omega^3+\omega^3 = 1+1+1 = \Gamma^{(1)}$. $\Gamma^{(2)}\Gamma^{(2)} = (1+\omega+\omega^2)\cdot(1+\omega+\omega^2) = 1+\omega^2+\omega = \Gamma^{(3)}$. Of course $\Gamma^{(1)}\Gamma^{(2)} = \Gamma^{(2)}$. Either $\Gamma^{(2)}$ or $\Gamma^{(3)}$ generate the rest of the group (which is isomorphic to A_3). A_3 is abelian moreover, hence it is ***cyclic***. When we used the character table of A_3 to build invariant functions, we didn't lose any elements of A_3.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{bmatrix}\begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} s_1+s_2+s_3 \\ s_1+\omega s_2+\omega^2 s_3 \\ s_1+\omega^2 s_2+\omega s_3 \end{bmatrix}$$

On the other hand A_5 is not cyclic because it is nonabelian. Its character table has 0s. When we multiply its character by a column vector of roots to make invariant functions from the roots we lose terms

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 0 & \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} & -1 \\ 4 & 1 & -1 & -1 & 0 \\ 3 & 0 & \frac{1-\sqrt{5}}{2} & \frac{1+\sqrt{5}}{2} & -1 \\ 5 & -1 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix} = \begin{bmatrix} r_1+r_2+r_3+r_4+r_5 \\ 3r_1+0\cdot r_2+\frac{1+\sqrt{5}}{2}r_3+\frac{1-\sqrt{5}}{2}r_4-r_5 \\ 4r_1+r_2-r_3-4r_4+0\cdot r_5 \\ 3r_1+0\cdot r_2+\frac{1-\sqrt{5}}{2}r_3+\frac{1+\sqrt{5}}{2}r_4-r_5 \\ 5r_1-r_2+0\cdot r_3+0\cdot r_4+r_5 \end{bmatrix}$$

The first linear combination involves all of the five roots, but none of the other combinations contains all five roots. So are missing more than one root. To see a solution to the quintic by radical extension, we need to construct functions MADE FROM the second, third, fourth and fifth equations above that are invariant under the interchange of symbols. Impossible. Hence the quintic is unsolvable by radical extension.

Be water my friend. “One step beyond the solvable equation,” Sander Bessels, PhD dissertation 2006, uses Galois theory, Klein’s icosahedron and elliptic curves to solve the general quintic. To the symmetries of the Standard Model, physicists have added supersymmetry (SUSY). Many infinites have died. Let’s proceed to the last of the algebraic blocks. After dealing with discrete groups, Jones proceeds to continuous groups. These notes provide several step-by-step examples that detail the theory developed in Jones and, supposedly, in H. Georgi’s book on Lie algebras in particle physics from Isospin to unified theories.

Self-contained algebra prerequisites for the Standard Model, GUTs and TOEs. We will study SU(2) and SU(3) in detail via tensor product and direct sum matrix mathematics, ladder operators and Young’s tableaux. This is where your background in junior or senior level exposure to quantum mechanical addition of angular momentum will come in handy. It is your option whether you read the material on the Wigner-Eckart theorem, which is useful for computing selection rules. What we learn from SU(2) and SU(3) will then be generalized to study the Cartan subalgebra, the attendant ladder operators, the Killing form, root and weight diagrams and Dynkin diagrams. These are the tools to “play” group theory with theories beyond the Standard Model, *e.g.*, SU(5) or Lissi’s TOE.

Basics [Highly augmented with my notes the following material is excerpted from “Spin and the addition of angular momentum using tensor notation,” Joel C. Corbo, 2007]. The aim of this section is to see the explicit connection between the coupled and uncoupled basis (between tensor products and direct sums that is). The spin operators S_x, S_y, and S_z have commutation relations given by

$$[S_x, S_y] = i\hbar S_z, \qquad [S_y, S_z] = i\hbar S_x, \qquad [S_z, S_x] = i\hbar S_y.$$

The above also holds for orbital angular momentum, but unlike spin, orbital angular momentum only takes integer values. Pauli proposed the following matrices with the same commutator algebra:

$$S_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad S_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad S_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

You may verify the commutation relations hold for these matrices. Let $S^2 = S_x^2 + S_y^2 + S_z^2$.

In matrix form we have

$$S^2 = \frac{3}{4}\hbar^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Verify S^2 matrix commutes with S_x, S_y, and S_z. Wrt the z axis let's define spin up and down states by

$$\left|\frac{1}{2}\ \frac{1}{2}\right\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \left|\frac{1}{2} - \frac{1}{2}\right\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

We see that the spin up eigenvector (or eigenstate) $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ has eigenvalue $\frac{1}{2}\hbar$ for the matrix S_z. The eigenvalue for the spin down eigenvector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ has eigenvalue $\frac{1}{2}\hbar$ for matrix S_z.

Please refer to a good textbook in quantum mechanics such as "Quantum Mechanics," Claude Cohen-Tannoudji, Bernard Diu and Franck Laloë, or "Modern Quantum Mechanics," JJ Sakurai for the required background to the following material on where matrices such as the Pauli matrices come from. The matrix algebra is very easy to follow. For the operator S_z let

$$S_z|s\ m_s\rangle = m_s\hbar|sm_s\rangle$$

where s can be a positive half integer or positive integer and m_s takes on all values between $-s$ and s. (Clearly we are generalizing beyond the spin ½ electron.) Then Pauli matrix S_z for, say, the electron is

$$\begin{pmatrix} \left\langle \frac{1}{2}\,\frac{1}{2}\middle| S_z\hbar \middle| \frac{1}{2}\,\frac{1}{2}\right\rangle & \left\langle \frac{1}{2}\,\frac{1}{2}\middle| S_z\hbar \middle| \frac{1}{2}\,-\frac{1}{2}\right\rangle \\ \left\langle \frac{1}{2}\,-\frac{1}{2}\middle| S_z\hbar \middle| \frac{1}{2}\,\frac{1}{2}\right\rangle & \left\langle \frac{1}{2}\,-\frac{1}{2}\middle| S_z\hbar \middle| \frac{1}{2}\,-\frac{1}{2}\right\rangle \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\hbar\left\langle \frac{1}{2}\,\frac{1}{2}\middle| \frac{1}{2}\,\frac{1}{2}\right\rangle & -\frac{1}{2}\hbar\left\langle \frac{1}{2}\,\frac{1}{2}\middle| \frac{1}{2}\,-\frac{1}{2}\right\rangle \\ \frac{1}{2}\hbar\left\langle \frac{1}{2}\,-\frac{1}{2}\middle| \frac{1}{2}\,\frac{1}{2}\right\rangle & -\frac{1}{2}\hbar\left\langle \frac{1}{2}\,-\frac{1}{2}\middle| \frac{1}{2}\,-\frac{1}{2}\right\rangle \end{pmatrix}$$

$$= \frac{1}{2}\hbar\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

If it happened that the particle in question was a spin one particle, then the matrix S_z would be

$$S_z = \frac{1}{2}\hbar\begin{pmatrix} \langle 1\,1|S_z|1\,1\rangle & \langle 1\,1|S_z|1\,0\rangle & \langle 1\,1|S_z|1-1\rangle \\ \langle 1\,0|S_z|1\,1\rangle & \langle 1\,0|S_z|1\,0\rangle & \langle 1\,0|S_z|1-1\rangle \\ \langle 1-1|S_z|1\,1\rangle & \langle 1-1|S_z|1\,0\rangle & \langle 1-1|S_z|1-1\rangle \end{pmatrix} = \frac{1}{2}\hbar\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let me introduce two new operators and corresponding matrices built from linear combinations of S_x and S_y. Let $S_\pm = S_x \pm iS_y$. By inversion, $S_x = \frac{1}{2}(S_+ + S_-)$ and $S_y = \frac{1}{2i}(S_+ - S_-)$. Check it out. Add S_x to iS_y to get $\frac{1}{2}S_+ + \frac{1}{2}S_- + \frac{1}{2}S_+ - \frac{1}{2}S_- = S_+$. For $S_\pm$ we have

$$S_\pm \left| s\, m_s \right\rangle = \hbar\sqrt{s(s+1) - m_s(m_s \pm 1)}|s m_s \pm 1\rangle.$$

For an electron let's compute the matrix for S_+.

$$\hbar\begin{pmatrix} \left\langle \frac{1}{2}\,\frac{1}{2}\middle| S_+ \middle| \frac{1}{2}\,\frac{1}{2}\right\rangle & \left\langle \frac{1}{2}\,\frac{1}{2}\middle| S_+ \middle| \frac{1}{2}\,-\frac{1}{2}\right\rangle \\ \left\langle \frac{1}{2}\,-\frac{1}{2}\middle| S_+ \middle| \frac{1}{2}\,\frac{1}{2}\right\rangle & \left\langle \frac{1}{2}\,-\frac{1}{2}\middle| S_+ \middle| \frac{1}{2}\,-\frac{1}{2}\right\rangle \end{pmatrix} = \hbar\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

This jives with summing the matrices for S_x and iS_y. Lastly for S^2 we have

$$S^2|s\, m_s\rangle = \hbar^2 s(s+1)|s\, m_s\rangle.$$

Let's proceed to describing systems with more than one particle. First let's restrict ourselves to two particles, one with spin s_1 and another with spin s_2. It turns out, as you shall see, that there are two bases for describing the possible spin states these two particles can be in. Let's consider the so-called uncoupled basis (the tensor product basis). We know that the first particle can be in one of $2s_1 + 1$ states and that the second particle can be in one of $2s_2 + 1$ states. If these particles are treated as independent, then there can be $(2s_1 + 1) \times (2s_2 + 1)$ possible states. This is denoted in tensor product notation thusly: $|s_1\, m_{s_1}\rangle \otimes |s_2\, m_{s_2}\rangle$. We are effectively forming a new Hilbert space for our particles' spins as a product of each individual particle's original Hilbert space. What does this mean? A few examples should clear this up. First let's work with operators and then with matrices. For example if the first particle is in a spin down state and the second particle is a spin up state, then $S_z^{(1)}$ operates only on the first particle

$$S_z^{(1)} \left|-\frac{1}{2}\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle = -\frac{1}{2}\hbar \left|-\frac{1}{2}\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle.$$

The operator only acts on the first particle's ket. Instead of working with operators, perhaps things will be more clear using matrices. Let

$$\left|\frac{1}{2}\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \quad \left|\frac{1}{2}\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}-\frac{1}{2}\right\rangle = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \quad \left|\frac{1}{2}-\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle = \begin{pmatrix}0\\0\\1\\0\end{pmatrix},$$

$$\left|\frac{1}{2}-\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}-\frac{1}{2}\right\rangle = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

represent the {spin up, spin up}, {spin up, spin down}, {spin down, spin up} and {spin down, spin down} states—our Hilbert state. Define the 4×4 matrix that only operates on the first particle's spin state as

$$S_z^{(1)} = S_z \otimes \mathbb{I}(2).$$

$\mathbb{I}(2)$ is the 2×2 identity matrix that leaves the second particle's spin state invariant. That is,

$$S_z^{(1)} = \frac{\hbar}{2}\begin{pmatrix} 1\cdot\mathbb{I}(2) & 0\cdot\mathbb{I}(2) \\ 0\cdot\mathbb{I}(2) & 1\cdot\mathbb{I}(2) \end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Then our example in matrix form with $S_z^{(1)}$ operating on the state with the first particle in the spin down state and the second particle in the spin up state is

$$\frac{\hbar}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\cdot\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = -\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

Note that if the first particle is a spin 1 particle it could be in one of three states along the arbitrary z-axis, namely in states 1, 0, -1, and if the second particle is a spin 3 particle it could be in one of seven states: 3, 2, 1, 0, -1, -2, -3. The matrix for $S_z^{(1)}$ would be an unwieldy 21×21 beast. The eigenvalues of $S_z^{(1)}$ are $\frac{\hbar}{2}\{1,1,-1,-1\}$. The eigenvectors are the columns of the 4×4 matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Let's stick to two spin ½ particles for now and build the remaining matrices for the tensor product (uncoupled) basis. What does $S_z^{(2)}$ look like? $S_z^{(2)} = \mathbb{I}(1) \otimes S_z^{(2)}$. In matrix form this would look like

$$S_z^{(2)} = \frac{\hbar}{2}\begin{pmatrix} 1\cdot S_z^{(2)} & 0\cdot S_z^{(2)} \\ 0\cdot S_z^{(2)} & 1\cdot S_z^{(2)} \end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

This matrix has eigenvalues $\frac{\hbar}{2}\{1,-1,1,-1\}$ with the same eigenvectors as $S_z^{(1)}$. Combining both matrices $S_z^{(1)}$ and $S_z^{(2)}$ we get

$$S_z = \hbar\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & -1\end{pmatrix}.$$

The eigenvalues of S_z are $\hbar\{1,0,0,-1\}$. It has the same eigenvectors as $S_z^{(1)}$ and $S_z^{(2)}$. This matrix operator, put together by tensor products, operates on the combined Hilbert space of the two electrons. What does S_z do to the state with both spins up?

$$\hbar\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & -1\end{pmatrix}\cdot\begin{pmatrix}1\\0\\0\\0\end{pmatrix} = \hbar\cdot\begin{pmatrix}1\\0\\0\\0\end{pmatrix}.$$

The spin is 1 and it is up. What about a system with the first particle spin up and the second particle spin down?

$$\hbar\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & -1\end{pmatrix}\cdot\begin{pmatrix}0\\1\\0\\0\end{pmatrix} = \hbar\cdot 0\cdot\begin{pmatrix}0\\1\\0\\0\end{pmatrix}.$$

The spin relative to our arbitrary z-axis is zero. Let's proceed to building the matrices for the remaining two axes. First, in tensor product form we have $S_x^{(1)} = S_x(1)\otimes\mathbb{I}(2)$. In matrix form this becomes

$$S_x^{(1)} = \frac{\hbar}{2}\begin{pmatrix}0\cdot\mathbb{I}(2) & 1\cdot\mathbb{I}(2)\\ 1\cdot\mathbb{I}(2) & 0\cdot\mathbb{I}(2)\end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix}0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{pmatrix}.$$

Then $S_x^{(2)} = \mathbb{I}(1)\otimes S_x(2)$ looks like

$$S_x^{(2)} = \frac{\hbar}{2}\begin{pmatrix}1\cdot S_x(2) & 0\cdot S_x(2)\\ 0\cdot S_x(2) & 1\cdot S_x(2)\end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix}0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix}.$$

Lastly $S_y^{(1)} = S_y(1)\otimes\mathbb{I}(2)$ in matrix form is

$$S_y^{(1)} = \frac{\hbar}{2}\begin{pmatrix} 0 \cdot \mathbb{I}(2) & -i \cdot \mathbb{I}(2) \\ -i \cdot \mathbb{I}(2) & 0 \cdot \mathbb{I}(2) \end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \\ -i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix}$$

and $S_y^{(2)} = \mathbb{I}(1) \otimes S_y(2)$ in matrix form is

$$S_x^{(2)} = \frac{\hbar}{2}\begin{pmatrix} 1 \cdot S_y(2) & 0 \cdot S_y(2) \\ 0 \cdot S_y(2) & 1 \cdot S_y(2) \end{pmatrix} = \frac{\hbar}{2}\begin{pmatrix} 0 & -i & 0 & 0 \\ -i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & -i & 0 \end{pmatrix}.$$

Just as we could define vectors in three dimensional Euclidean space with three unit basis vectors $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$, e.g., $\vec{\mathbf{A}} = 4\hat{\mathbf{i}} - 2\hat{\mathbf{j}} + \frac{1}{2}\hat{\mathbf{k}}$, the spin state of particle one has three basis vectors $S_x^{(1)}, S_y^{(1)}, S_z^{(1)}$. The basis vectors for the spin one particle happen to be matrices. The same applies for particle 2. Let's practice some of this vector algebra. $S_1^2 = \mathbf{S}_1 \cdot \mathbf{S}_1 = S_x^2 + S_y^2 + S_z^2$. In matrix form this is

$$S_1^2 = \frac{\hbar^2}{4}\left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right] = \frac{\hbar^2}{4}\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

We get the same result for S_2^2. We can compute the dot product $\boldsymbol{S}_1 \cdot \mathbf{S}_2 = S_{1x} \cdot S_{2x} + S_{1y} \cdot S_{2y} + S_{1z} \cdot S_{2z}$. In matrix form this is

$$\boldsymbol{S}_1 \cdot \mathbf{S}_2 = \frac{\hbar^2}{4}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

This allows us to compute $S^2 = (\mathbf{S}_1 + \mathbf{S}_2)^2$. Note $S^2 = (\mathbf{S}_1 + \mathbf{S}_2)^2 = S_1^2 + S_2^2 + 2\boldsymbol{S}_1 \cdot \mathbf{S}_2$. Therefore

$$S^2 = \hbar^2\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

Here at last we will start to see the connection between the uncoupled (tensor) product basis and the direct sum basis which I have yet to introduce. The eigenvalues of S^2 are $\hbar^2\{2,2,0,2\}$. The corresponding eigenvectors are the columns of the matrix

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In terms of {spin up, spin up}, {spin up, spin down}, {spin down, spin up} and {spin down, spin down} states

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

we have a set of "coupled" (direct sum) states (to be explained immediately following the labeling) $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

for the {spin up, spin up} state, and $\frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ a coupled mix of the {spin up, spin down} state with the {spin down, spin up} state. We have another possible coupled mix of the {spin up, spin down} state with the {spin down, spin up} state

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

and we have $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ for the {spin down, spin down} state. The {spin up, spin up} and {spin down, spin down} states in the coupled (direct sum) basis are the same as before in the uncoupled basis, but not so the {spin up, spin down} and {spin down, spin down} states. In physics the coupled (direct sum) basis we have arrived at arises when the Hamiltonian includes the "spin-spin" interaction $\boldsymbol{S}_1 \cdot \mathbf{S}_2$. In the uncoupled, tensor product basis the basis vectors are

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

In the coupled, direct sum basis the basis vectors are

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

They are connected by

$$U^\dagger S^2 U = 2\hbar^2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Do you see Schrödinger's cat? When the two electrons are coupled by the dot product, two states arise that are superposed from the pure states $\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$.

Let's formalize the coupled, direct sum basis. In this basis we emphasize the composite nature of a system with a particle of spin s_1 with $2s_1 + 1$ possible spin states in the z-direction and s_2 with

$2s_2+1$ possible spin states in the z direction. If, say, spin s_2 was larger than s_1 then counting the number of states along the z-direction would tabulate like:

$$\begin{bmatrix} s_1+s_2 & (s_1-1)+s_2 & \cdots & (s_1-2s_1)+s_2 \\ s_1+(s_2-1) & (s_1-1)+(s_2-1) & & (s_1-2s_1)+(s_2-1) \\ \vdots & & \ddots & \vdots \\ s_1+(s_2-(2s_2-1)) & (s_1-1)+\left(s_2-(2s_2-1)\right) & \cdots & (s_1-2s_1)+(s_2-(2s_2-1)) \\ s_1+(s_2-2s_2) & (s_1-1)+(s_2-2s_2) & & (s_1-2s_1)+(s_2-2s_2) \end{bmatrix}.$$

Cleaning this counting of the $(2s_1+1)\times(2s_2+1)$ states along the z axis we can get

$$\begin{bmatrix} s_1+s_2 & (s_1+s_2)-1 & \cdots & s_2-s_1 \\ (s_1+s_2)-1 & (s_1+s_2)-2 & & (s_2-s_1)-1 \\ \vdots & & \ddots & \vdots \\ (s_1-s_2)+1 & s_1-s_2 & \cdots & -(s_1+s_2)+1 \\ s_1-s_2 & (s_1-s_2)-1 & & -(s_1+s_2) \end{bmatrix}$$

Each entry in this table represents the total spin in the z-direction for some state of our two particle system. But what if we didn't know about the fact that we had two particles, and we simply think of the table as corresponding to the spin of a system. The red entries run from s_1+s_2 and $-(s_1+s_2)$ in steps of 1. Likewise the blue entries would run from s_1+s_2-1 through$-(s_1+s_2-1)$, and so on until we get to the green entries which run from s_2-s_1 through $-(s_2-s_1)$. Hence we'd conclude that this table is a table of all possible spin states associated with spin $s=\{s_1+s_2, s_1+s_2-1, \ldots, |s_2-s_1|\}$.

Since we do in fact know that we constructed this table by summing the z-components of the spin of two particles, we conclude that the Hilbert space of our two particles of spins s_1 and s_2 can also be constructed out of all states represented by $s=\{s_1+s_2, s_1+s_2-1, \ldots, |s_2-s_1|\}$. Symbolically we write our "coupled" Hilbert space as

$$\left|s_1+s_2\; m_{s_1+s_2}\right\rangle \oplus \left|s_2+s_1-1\; m_{s_1+s_2-1}\right\rangle \oplus \ldots \oplus \left||s_2-s_1| m_{|s_2-s_1|}\right\rangle,$$

where $\oplus$ operation is called a direct sum.

Now, putting it all together, the theory with our two electron example, we have

$$\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0.$$

The uncoupled Hilbert space of two electrons (two spin ½ particles) with basis

$$\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \quad \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \quad \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \quad \begin{pmatrix}0\\0\\0\\1\end{pmatrix}.$$

can be transformed to the direct sum coupled Hilbert space with three symmetric states and one antisymmetric state. Let me explain this. Let A represent the first electron and B represent the second electron. In the coupled basis the first state is (spin up, spin up), or A, B. Since electrons are indistinguishable, I could swap labels A and B and not change anything. The second state (spin up, spin down) is $\frac{1}{\sqrt{2}}(A + B)$. If I swap labels I get $\frac{1}{\sqrt{2}}(B + A)$. Skipping to the fourth state, (spin down, spin down) we have A, B, and swapping labels doesn't change anything. For the third state something changes. The third state is again a combination of (spin up, spin down), but in terms of A and B it is $\frac{1}{\sqrt{2}}(A - B)$. If I swap A with B, you get $\frac{1}{\sqrt{2}}(B - A) = -\frac{1}{\sqrt{2}}(A - B)$. This state is antisymmetric under the interchange of A and B while the other states are symmetric under the interchange of A and B. We have a triplet of symmetric states and a singlet of one antisymmetric state. In an incomplete model of physics can think of mesons as being composed of three spin ½ quarks and nucleons as being composed of three spin ½ quarks. (It's better to get the history of the strong nuclear force, mesons, and quarks from a book on particle physics like the one by Griffiths.) In this crude model of physics the pion spectrum (made from two spin ½ quarks) would be explained by $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$, and nucleons would be explained by $\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} = \frac{3}{2} \oplus \frac{1}{2} \oplus \frac{1}{2}$. We can build up these kinds of particle spectra using the matrix methods we have been playing with, but there are two alternative methods as well: via ladder operators and via Young's tableaux. Let's mop up the matrix language for tensor products and direct sums.

Matrix language for tensor products. Let me formalize the matrices we have formed in our studies of tensor product spaces. Suppose we have two 2×2 matrices $\mathbf{A}$ and $\mathbf{B}$ defined by

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

The tensor product of $\mathbf{A}$ and $\mathbf{B}$ is given by

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} A_{11}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} & A_{12}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \\ A_{21}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} & A_{22}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} & A_{11}B_{12} & A_{12}B_{11} & A_{12}B_{12} \\ A_{11}B_{21} & A_{11}B_{22} & A_{12}B_{21} & A_{12}B_{22} \\ A_{21}B_{11} & A_{21}B_{12} & A_{22}B_{11} & A_{22}B_{12} \\ A_{21}B_{21} & A_{21}B_{22} & A_{22}B_{21} & A_{22}B_{22} \end{pmatrix}.$$

Recall that we began with the following assignments: $\left|\frac{1}{2}\,\frac{1}{2}\right\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\left|\frac{1}{2} - \frac{1}{2}\right\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$? Here is how we got our tensor product states which I had previously just told you:

$$\left|\frac{1}{2}\,\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}\,\frac{1}{2}\right\rangle = \begin{pmatrix} 1\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ 0\begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \left|\frac{1}{2} - \frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}\,\frac{1}{2}\right\rangle = \begin{pmatrix} 0\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ 1\begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

$$\left|\frac{1}{2}\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2} - \frac{1}{2}\right\rangle = \begin{pmatrix} 1\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ 0\begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \text{and} \quad \left|\frac{1}{2} - \frac{1}{2}\right\rangle \otimes \left|\frac{1}{2} - \frac{1}{2}\right\rangle = \begin{pmatrix} 0\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ 1\begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

When we computed S_z in matrix form we were computing $S_z = S_z(1) \otimes \mathbb{I}(2) + \mathbb{I}(1) \otimes S_z(2)$.

Direct sums of matrices in terms of matrices. If $\mathbf{A}$ and $\mathbf{B}$ are matrices of dimension $m \times n$ and $p \times q$ respectively, the direct sum is defined by:

$$\mathbf{A} \oplus \mathbf{B} = \begin{pmatrix} A_{m\times n} & 0_{m\times q} \\ 0_{p\times q} & B_{p\times q} \end{pmatrix}$$

where the $\mathbf{0}$ matrix is a matrix of zeroes. Recall from our two electron example that we arrived at $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$ where on the LHS (tensor product side of two spin ½ electrons) we had the {spin up, spin up}, {spin up, spin down}, {spin down, spin up} and {spin down, spin down} states given by:

$$\left|\frac{1}{2}\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \quad \left|\frac{1}{2}\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}-\frac{1}{2}\right\rangle = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \quad \left|\frac{1}{2}-\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle = \begin{pmatrix}0\\0\\1\\0\end{pmatrix},$$

$$\left|\frac{1}{2}-\frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}-\frac{1}{2}\right\rangle = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

Recall also that the red states corresponded to a triplet of symmetric states, and that the sole blue state corresponded to an antisymmetric singlet. On the RHS, the side with $1 \oplus 0$ it appears that we have a spin 1 particle coupled with a spin 0 particle. The spin 1 particle has Hilbert space $|1\ m_{s=1} = 1\rangle$, $|1\ m_{s=1} = 0\rangle$, and $|1\ m_{s=1} = -1\rangle$. The spin 0 particle has Hilbert space $|0\ m_{s=0} = 0\rangle$. From this point of view another assignment of the spin states makes sense. Let

$$|1\ 1\rangle = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \quad |1\ 0\rangle = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \quad |1\ -1\rangle = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \quad |0\ 0\rangle = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}.$$

In this **<u>convention</u>** we list the states in decreasing order by s, and within each s multiplet, in descending order by m_s. In the coupled, $\oplus$ basis then

$$S_z = S_{z,s=1} \oplus s_{z,s=0} = \hbar \begin{pmatrix}1 & 0 & 0\\0 & 0 & 0\\0 & 0 & -1\end{pmatrix} \oplus \hbar(0) = \hbar \begin{pmatrix}1 & 0 & 0 & 0\\0 & 0 & 0 & 0\\0 & 0 & -1 & 0\\0 & 0 & 0 & 0\end{pmatrix}.$$

This matrix s_z of a spin 1 particle coupled by direct sum $\oplus$ to a spin 0 particle has the same eigenvectors and eigenvalues as the picture of two uncoupled spin ½ particles under tensor product $\otimes$ with matrix

$$S_z = \hbar \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

I could align these two different pictures (two spin ½ particles under tensor product) with one spin 1 particle and 1 spin 0 particle under direct sum by turning the label to

$$\left|\frac{1}{2} - \frac{1}{2}\right\rangle \otimes \left|\frac{1}{2} - \frac{1}{2}\right\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

and

$$\left|\frac{1}{2} - \frac{1}{2}\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

The Hilbert space cannot distinguish if you have two spin ½ particles under direct product, or a spin 1 particle with a spin 0 particle under direct sum. You the physicist know this. A system consisting of a spin s_1 particle and a spin s_2 particle, under the uncoupled tensor product representation live in a Hilbert space given by $|s_1 m_{s_1}\rangle \otimes |s_2 m_{s_2}\rangle$. In the coupled direct sum representation (using our updated convention for the basis states) live in a Hilbert space given by

$$|s_1 + s_2 m_{s_1+s_2}\rangle \oplus |s_1 + s_2 m_{s_1+s_2-1}\rangle \oplus \dots \oplus \big||s_2 - s_1|\, m_{|s_2-s_1|}\big\rangle.$$

Since these two Hilbert spaces are equivalent, it must be that

$$|s_1 m_{s_1}\rangle \otimes |s_2 m_{s_2}\rangle = |s_1 + s_2 m_{s_1+s_2}\rangle \oplus |s_1 + s_2 m_{s_1+s_2-1}\rangle \oplus \dots \oplus \big||s_2 - s_1|\, m_{|s_2-s_1|}\big\rangle.$$

This is the addition of **<u>angular momentum</u>** in quantum mechanics. It applies to the addition of angular momentum for both spin and orbital angular momentum. We can equally well represent the system in terms of eigenstates of the individual particle's angular momentum operators, or in terms of the system's total angular momentum operators. We first derived $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$ the long way. Now that we have the mathematics better identified between the uncoupled and coupled bases, we have

$$s_1 \otimes s_2 = s_1 + s_2 \oplus s_1 + s_2 - 1 \oplus \dots \oplus |s_2 - s_1|,$$

so

$$\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0, \qquad \text{and say,} \qquad \frac{3}{2} \otimes 2 = \frac{7}{2} \oplus \frac{5}{2} \oplus \frac{3}{2} \oplus \frac{1}{2}.$$

Here is a useful observation. Given this notation, note that if we replace the quantum number representing a state with the multiplicity of the state and change the operators $\otimes$ to $\times$ and $\oplus$ to $+$, we should always get a correct equation. Take $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$ for example. Each spin $\frac{1}{2}$ state on the LHS has $m_s = \left\{\frac{1}{2}, -\frac{1}{2}\right\}$. Thus we may replace $\frac{1}{2} \otimes \frac{1}{2}$ with 2×2. On the RHS we have spin 1 with $m_s = \{1, 0, -1\}$ and spin 0 with only $m_s = 0$. Thus on the RHS we may replace $1 \oplus 0$ with $3 + 1$. So $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$ translates to $2 \times 2 = 3 + 1$. The state with spin $\frac{3}{2}$ has $m_s = \left\{\frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}\right\}$ and similarly the state with spin 2 has $m_s = \{2, 1, 0, -1, -2\}$. Thus there are a total of 20 basis states in $\frac{3}{2} \otimes 2$. We know that spin $\frac{1}{2}$ corresponds to two states. We know that spin $\frac{3}{2}$ corresponds to 4 states. Spin $\frac{5}{2}$ has $m_s = \left\{\frac{5}{2}, \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}, -\frac{5}{2}\right\}$ values. There are 6 of these. Finally spin $\frac{7}{2}$ has 8 m_s possibilities. So

$$\frac{3}{2} \otimes 2 = \frac{7}{2} \oplus \frac{5}{2} \oplus \frac{3}{2} \oplus \frac{1}{2}$$

translates to $4 \times 5 = 8 + 6 + 4 + 2$. This works because we're counting the total number of basis states in both representations of our Hilbert space.

How do the basis states in the uncoupled tensor product state look like in the coupled direct sum basis states or conversely? We're going to get to this presently, but for completeness I will also talk about Clebsch-Gordon coefficients and show the method of Young's tableaux. All of this material will be developed by example. Let's begin with a spin 1 and spin $\frac{1}{2}$ particle. In the uncoupled tensor product representation the basis states are

$$|11\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle, \qquad |10\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle, \qquad |1{-}1\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle,$$

$$|11\rangle \otimes \left|\frac{1}{2}-\frac{1}{2}\right\rangle, \qquad |10\rangle \otimes \left|\frac{1}{2}-\frac{1}{2}\right\rangle, \qquad |1{-}1\rangle \otimes \left|\frac{1}{2}-\frac{1}{2}\right\rangle.$$

From $s_1 \otimes s_2 = s_1 + s_2 \oplus s_1 + s_2 - 1 \oplus \dots \oplus |s_2 - s_1|$ we know $1 \otimes \frac{1}{2} = \frac{3}{2} \oplus \frac{1}{2}$. In the coupled direct sum space the basis vectors are

$$\left|\frac{3}{2}\frac{3}{2}\right\rangle, \qquad \left|\frac{3}{2}\frac{1}{2}\right\rangle, \qquad \left|\frac{3}{2}-\frac{1}{2}\right\rangle, \qquad \left|\frac{3}{2}-\frac{3}{2}\right\rangle,$$

$$\left|\frac{1}{2}\frac{1}{2}\right\rangle, \qquad \left|\frac{1}{2}-\frac{1}{2}\right\rangle.$$

Let's start matching the coupled states to linear combinations of the uncoupled states using the fact that both representations are eigenstates of the total operator S_z and the orthogonality of the basis states in either representation. To begin, there is one state in each representation with $S_z = \frac{3}{2}\hbar$, namely $\left|\frac{3}{2}\frac{3}{2}\right\rangle = |11\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle$. This is our first linear combination with only one term on each side of the equation. There is another easy one, namely $\left|\frac{3}{2}-\frac{3}{2}\right\rangle = |1{-}1\rangle \otimes \left|\frac{1}{2}-\frac{1}{2}\right\rangle$.

Now we run into problems because there are two states in each representation that have $S_z = \frac{1}{2}\hbar$. These are $\left|\frac{3}{2}\frac{3}{2}\right\rangle$ and $\left|\frac{3}{2}\frac{1}{2}\right\rangle$. We will use the lowering operator (in matrix element form) with the negative sign

$$S_{\pm}\left|s\, m_s\right\rangle = \hbar\sqrt{s(s+1) - m_s(m_s \pm 1)}\left|s m_s \pm 1\right\rangle.$$

In operator form this is $S_- = S_-(1) \otimes \mathbb{I}(2) + \mathbb{I}(1) \otimes S_-(2)$. So

$$S_-\left|\frac{3}{2}\frac{3}{2}\right\rangle = \left(S_-(1) \otimes \mathbb{I}(2) + \mathbb{I}(1) \otimes S_-(2)\right)\left|11\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle.$$

$$\hbar\sqrt{\frac{3}{2}\left(\frac{3}{2}+1\right) - \frac{3}{2}\left(\frac{3}{2}-1\right)}\left|\frac{3}{2}\frac{1}{2}\right\rangle = S_-(1) \otimes \mathbb{I}(2)\left|11\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle + \mathbb{I}(1) \otimes S_-(2)\left|\frac{1}{2}\frac{1}{2}\right\rangle \otimes \left|11\right\rangle$$

Operator $S_-(1) \otimes \mathbb{I}(2)$ acts on the first ket; $\mathbb{I}(1) \otimes S_-(2)$ operates on the second ket. Proceeding:

$$\sqrt{3}\hbar\left|\frac{3}{2}\frac{1}{2}\right\rangle = \hbar\sqrt{1(1+1) - 1(1-1)}\left|10\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle + \left|11\right\rangle \otimes \hbar\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right) - \frac{1}{2}\left(\frac{1}{2}-1\right)}\left|\frac{1}{2}\frac{1}{2}\right\rangle$$

$$= \hbar\sqrt{2}\left|10\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle + \left|11\right\rangle \otimes \hbar\left|\frac{1}{2}\frac{1}{2}\right\rangle.$$

So the linear combination for $S_-\left|\frac{3}{2}\frac{1}{2}\right\rangle$ in the $\otimes$ basis is

$$S_-\left|\frac{3}{2}\frac{3}{2}\right\rangle = \left|\frac{3}{2}\frac{1}{2}\right\rangle = \frac{1}{\sqrt{3}}\left(S_-(1) \otimes \mathbb{I}(2) + \mathbb{I}(1) \otimes S_-(2)\right)\left|11\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle$$

$$= \sqrt{\frac{2}{3}}\left|10\right\rangle \otimes \left|\frac{1}{2}\frac{1}{2}\right\rangle + \sqrt{\frac{1}{3}}\left|11\right\rangle \otimes \left|\frac{1}{2}-\frac{1}{2}\right\rangle$$

Now we seek out the linear combination of for $\left|\frac{3}{2}\frac{1}{2}\right\rangle$ by the same route—let me do the LHS algebra first.

$$S_-\left|\frac{3}{2}\frac{1}{2}\right\rangle = \hbar\sqrt{\frac{3}{2}\left(\frac{3}{2}+1\right)-\frac{1}{2}\left(\frac{1}{2}-1\right)}\left|\frac{3}{2}-\frac{1}{2}\right\rangle = 2\hbar\left|\frac{3}{2}-\frac{1}{2}\right\rangle =$$

$$= (S_-(1)\otimes\mathbb{I}(2)+\mathbb{I}(1)\otimes S_-(2))\left(\sqrt{\frac{2}{3}}|10\rangle\otimes\left|\frac{1}{2}\frac{1}{2}\right\rangle+\sqrt{\frac{1}{3}}|11\rangle\otimes\left|\frac{1}{2}-\frac{1}{2}\right\rangle\right).$$

$$2\hbar\left|\frac{3}{2}-\frac{1}{2}\right\rangle = \sqrt{\frac{2}{3}}\,\hbar\sqrt{1(1+1)-0(0-1)}|1-1\rangle\otimes\left|\frac{1}{2}\frac{1}{2}\right\rangle+\hbar\sqrt{\frac{1}{3}}\sqrt{1(1+1)-1(1-1)}|10\rangle\otimes\left|\frac{1}{2}-\frac{1}{2}\right\rangle$$

$$+\hbar\sqrt{\frac{2}{3}}\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)-\frac{1}{2}\left(\frac{1}{2}-1\right)}|10\rangle$$

$$\otimes\left|\frac{1}{2}-\frac{1}{2}\right\rangle+\hbar\sqrt{\frac{1}{3}}\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)+\frac{1}{2}\left(-\frac{1}{2}-1\right)}|11\rangle\otimes\left|\frac{1}{2}-\frac{3}{2}\right\rangle.$$

$$2\left|\frac{3}{2}-\frac{1}{2}\right\rangle = \sqrt{\frac{2}{3}}\,\sqrt{2}|1-1\rangle\otimes\left|\frac{1}{2}\frac{1}{2}\right\rangle+\sqrt{2}\sqrt{\frac{1}{3}}|10\rangle\otimes\left|\frac{1}{2}-\frac{1}{2}\right\rangle+\sqrt{\frac{2}{3}}|10\rangle\otimes\left|\frac{1}{2}-\frac{1}{2}\right\rangle+0\cdot|11\rangle\otimes\left|\frac{1}{2}-\frac{3}{2}\right\rangle.$$

$$\left|\frac{3}{2}-\frac{1}{2}\right\rangle = \sqrt{\frac{1}{3}}\,|1-1\rangle\otimes\left|\frac{1}{2}\frac{1}{2}\right\rangle+\sqrt{\frac{1}{6}}|10\rangle\otimes\left|\frac{1}{2}-\frac{1}{2}\right\rangle+\sqrt{\frac{1}{6}}|10\rangle\otimes\left|\frac{1}{2}-\frac{1}{2}\right\rangle.$$

$$\left|\frac{3}{2}-\frac{1}{2}\right\rangle = \sqrt{\frac{1}{3}}\,|1-1\rangle\otimes\left|\frac{1}{2}\frac{1}{2}\right\rangle+\sqrt{\frac{2}{3}}|10\rangle\otimes\left|\frac{1}{2}-\frac{1}{2}\right\rangle.$$

We have found four of the coupled states in terms of linear combination of the coupled states. We already what lowering $\left|\frac{3}{2}-\frac{1}{2}\right\rangle$ will lead to, namely, $\left|\frac{3}{2}-\frac{3}{2}\right\rangle = |1-1\rangle\otimes\left|\frac{1}{2}-\frac{1}{2}\right\rangle$. We need to figure out how to compute the linear combinations for $\left|\frac{1}{2}\frac{1}{2}\right\rangle$, $\left|\frac{1}{2}-\frac{1}{2}\right\rangle$. It is time to use orthogonality. Suppose that

$$\left|\frac{1}{2}\frac{1}{2}\right\rangle = A|10\rangle\otimes\left|\frac{1}{2}\frac{1}{2}\right\rangle+B|11\rangle\otimes\left|\frac{1}{2}-\frac{1}{2}\right\rangle,$$

where A and B are to be determined. We know that $\left\langle \frac{3}{2}\frac{1}{2} \middle| \frac{1}{2}\frac{1}{2} \right\rangle = 0$.

$$\left\langle \frac{3}{2}\frac{1}{2} \middle| \frac{1}{2}\frac{1}{2} \right\rangle = \left\langle \sqrt{\frac{2}{3}}\langle 10| \otimes \left\langle \frac{1}{2}\ \frac{1}{2} \right| + \sqrt{\frac{1}{3}}\langle 11| \otimes \left\langle \frac{1}{2}\ -\frac{1}{2} \right| \middle| A|10\rangle \otimes \left| \frac{1}{2}\frac{1}{2} \right\rangle + B|11\rangle \otimes \left| \frac{1}{2} - \frac{1}{2} \right\rangle \right\rangle = 0.$$

This leads to $\sqrt{\frac{2}{3}}A + \sqrt{\frac{1}{3}}B = 0$. Also orthonormality requires $\left\langle \frac{1}{2}\frac{1}{2} \middle| \frac{1}{2}\frac{1}{2} \right\rangle$, leading to $|A|^{\mathbf{2}} + |B|^{\mathbf{2}} = 1$.

Hence $A = -\sqrt{\frac{1}{3}}$ and $B = \sqrt{\frac{2}{3}}$. Thus

$$\left| \frac{1}{2}\frac{1}{2} \right\rangle = -\sqrt{\frac{1}{3}}|10\rangle \otimes \left| \frac{1}{2}\frac{1}{2} \right\rangle + \sqrt{\frac{2}{3}}|11\rangle \otimes \left| \frac{1}{2} - \frac{1}{2} \right\rangle.$$

Applying the lower operator to $\left| \frac{1}{2}\frac{1}{2} \right\rangle$ leads to

$$\left| \frac{1}{2} - \frac{1}{2} \right\rangle = \sqrt{\frac{1}{3}}|10\rangle \otimes \left| \frac{1}{2}\frac{1}{2} \right\rangle - \sqrt{\frac{2}{3}}|11\rangle \otimes \left| \frac{1}{2} - \frac{1}{2} \right\rangle.$$

Note that the state $\left| \frac{1}{2} - \frac{1}{2} \right\rangle$ is not an eigenstates of $S_z^{(1)}$ nor $S_z^{(2)}$. This state is a mixture of states with different values of m_{s_1} and m_{s_2}. These results of writing coupled eigenstates in terms of linear combinations of uncoupled eigenstates is, of course, invertible. Seeing the uncoupled basis states as linear combinations of coupled eigenstates shows us that the uncoupled basis states are not eigenstates of S_{total}^2 because they are mixtures of states with different values of s_{total}.

Recall from above that

$$S_- \left| \frac{3}{2}\frac{3}{2} \right\rangle = \sqrt{\frac{2}{3}}|10\rangle \otimes \left| \frac{1}{2}\frac{1}{2} \right\rangle + \sqrt{\frac{1}{3}}|11\rangle \otimes \left| \frac{1}{2} - \frac{1}{2} \right\rangle$$

Note that these coefficients match the table of Clebsch-Gordan coefficients from the web site

http://en.wikipedia.org/wiki/Table_of_Clebsch%E2%80%93Gordan_coefficients . The table is generated

in terms of total angular momentum using the letter j instead of s.

j_1=1, j_2=1/2

m=3/2 j=

m$_1$, m$_2$=

	3/2
1, 1/2	1

m=1/2 j=

m$_1$, m$_2$=

	3/2	1/2
1, -1/2	$\sqrt{\frac{1}{3}}$	$\sqrt{\frac{2}{3}}$
0, 1/2	$\sqrt{\frac{2}{3}}$	$-\sqrt{\frac{1}{3}}$

What we derived piece by piece follows from the Clebsch-Gordan coefficients which are solutions to

$$|(j_1 j_2)jm\rangle = \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} |j_1 m_1 j_2 m_2\rangle\langle j_1 j_2; m_1 m_2|j_1 j_2; jm\rangle.$$

Explicitly:

$$\langle j_1 j_2; m_1 m_2|j_1 j_2; jm\rangle$$

$$= \sqrt{\frac{(2j+1)(j+j_1-j_2)!\,(J_1+j_2-j))!}{(j_1+j_2+j+1)!}}$$

$$\times \sqrt{(j+m)!\,(j-m)!\,(j_1+m_1)!\,(j_1+m_1)!\,(j_2+m_2)!\,(j_2-m_2)}$$

$$\times \sum_k \frac{(-1)^k}{k!\,(j_1+j_2-j-k)!\,(j_1-m_1-k)!\,(j_2+m_2-k)!\,(j-j_2+m_1+k)!\,(j-j_1-m_2+k)!},$$

where the summation is extended over all integer k for which the argument above every factorial is nonnegative. For brevity, solutions with $m < 0$ and $j_1 < j_2$ are omitted. They may be calculated using the relations

$$\langle j_1 j_2; m_1 m_2 | j_1 j_2; jm\rangle = (-1)^{j-j_1-j_2} \langle j_1 j_2; -m_1 - m_2 | j_1 j_2; j-m\rangle \text{ and}$$

$$\langle j_1 j_2; m_1 m_2 | j_1 j_2; jm\rangle = (-1)^{j-j_1-j_2} \langle j_1 j_2; m_1 m_2 | j_1 j_2; jm\rangle.$$

Let's pause to reflect a little about what we have been doing in terms of atoms with many electrons and in terms of quarks being constituent particles. The mathematics we have been enjoying so far have dealt with two spin ½ particles. We have learned that there is a triplet state of symmetric wave functions and a singlet state of antisymmetric wave functions. In historical nuclear physics the first two quarks were the up and down quark, and their antiparticles. The pions could then be described as members of an Isospin triplet: the π^+, π^0, π^- as follows

$$\pi^+ = \Big|11\rangle = \left|\frac{1}{2}\,\frac{1}{1}\right\rangle_1 \left|\frac{1}{2}\,\frac{1}{2}\right\rangle_2,$$

$$\pi^0 = |10\rangle = \frac{1}{\sqrt{2}}\left(\left|\frac{1}{2}\,\frac{1}{1}\right\rangle_1 \left|\frac{1}{2} - \frac{1}{2}\right\rangle_2 + \left|\frac{1}{2}\,\frac{1}{1}\right\rangle_1 \left|\frac{1}{2} - \frac{1}{2}\right\rangle_2\right),$$

$$\pi^- = \Big|1{-1}\rangle = \left|\frac{1}{2}\,\frac{1}{1}\right\rangle_1 \left|\frac{1}{2}\,\frac{1}{2}\right\rangle_2.$$

For the details in terms of quarks and antiquarks, please refer to chapter 12 by Melih Sener and Klaus Schulten in the PDF file at http://www.ks.uiuc.edu/Services/Class/PHYS480/qm_PDF/chp12.pdf .

Systems of more than two particles. Suppose we wanted to determine the "spectrum" of states one can conjure up with three spin ½ electrons or three quarks, the up, down and strange quarks for example. We would get

$$\frac{1}{2} \otimes \frac{1}{2} \otimes \frac{1}{2} = \left(\frac{1}{2} \otimes \frac{1}{2}\right) \otimes \frac{1}{2} = (1 \oplus 0) \otimes \frac{1}{2} = \left(\text{distribute the term} \frac{1}{2} \text{ to each term on the LHS}\right)$$

$$= \left(1 \otimes \frac{1}{2}\right) \oplus \left(0 \otimes \frac{1}{2}\right) = \left(\frac{3}{2} \oplus \frac{1}{2}\right) \oplus \frac{1}{2} = \frac{3}{2} \oplus \frac{1}{2} \oplus \frac{1}{2}.$$

We see that three spin $\frac{1}{2}$ particles combine to form a spin $\frac{3}{2}$ multiplet and two distinct spin $\frac{1}{2}$ multiplets. The difference between the two spin $\frac{1}{2}$ multiplets is symmetry: one of them is antisymmetric with respect to swapping the first two spin $\frac{1}{2}$ particles and the other symmetric with respect to that swapping.

A parting example:

$$1 \otimes 2 \otimes 3 = (1 \otimes 2) \otimes 3 = (3 \oplus 2 \oplus 1) \otimes 3 = (3 \otimes 3) \oplus (2 \otimes 3) \oplus (1 \otimes 3)$$

$$= (6 \oplus 5 \oplus 4 \oplus 3 \oplus 2 \oplus 1 \oplus 0) \oplus (5 \oplus 4 \oplus 3 \oplus 2 \oplus 1) \oplus (4 \oplus 3 \oplus 2)$$

$$= 6 \oplus 5 \oplus 5 \oplus 4 \oplus 4 \oplus 4 \oplus 3 \oplus 3 \oplus 3 \oplus 2 \oplus 2 \oplus 2 \oplus 1 \oplus 1 \oplus 0.$$

You may have some questions about this result, such as why $(3 \otimes 3)$ counts down to 0, but $(2 \otimes 3)$ only goes down to 1, and why $(1 \otimes 3)$ only goes down to 2. Why? Recall that

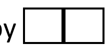
$$s_1 \otimes s_2 = s_1 + s_2 \oplus s_1 + s_2 - 1 \oplus \ldots \oplus |s_2 - s_1|.$$

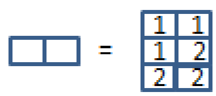
It is this result that determines that $(3 \otimes 3)$ counts down to 0, and that $(2 \otimes 3)$ only goes down to, and that $(1 \otimes 3)$ only goes down to 2. It's good to ask questions.

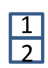
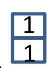
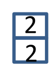
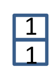
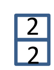
Young's tableaux. There is nothing wrong with the parting example above, but there is a graphical method for converting tensor products into direct sums that also does the job of collecting terms into one of three bins, these being totally symmetric states, totally antisymmetric states, and states with mixed symmetry. This is not implicit in the method shown in the parting example above.

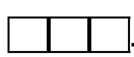
The graphical method is called the method of Young's tableaux. I follow the development of this tool from JJ Sakurai's book on modern quantum mechanics. All of these (SU(2) and (SU(3)) intricacies will generalize to a fairly straightforward set of general tools that underpin the study of Lie groups, such as, say SU(5,) the basis of the first Grand Unified Theory (GUT) or Lissi's Theory of Everything (TOE), E8.

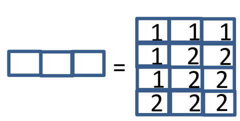
The spin state of an individual electron is to be represented by a box. We let [1] represent spin up and [2] represent spin down. These boxes are the primitives of SU(2), with a single box representing a doublet.

We define a symmetric tableau by  (think of the spin triplet of two electrons) and an antisymmetric tableau by , the spin singlet of two electrons. That is

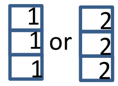


We don't consider [2][1] because when boxes are placed horizontally, symmetry is understood. Double counting is avoided if we require that the number not decrease going from the left to the right. As for the singlet  there is only one possibility: . You cannot have  nor  since there is no way to make these tableaux antisymmetric. To avoid double counting we do not write  or . To eliminate the unwanted symmetry states we require the number to increase as we go down.

Now consider three electrons . We can construct totally symmetric spin states by the following

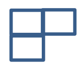
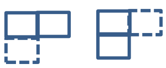


Each lower row has increased from the row above it. The four eigenstates are the multiplicity of the $j = \frac{3}{2}$ state. That it is symmetric can be seen for the $m = \frac{3}{2}$ case where all three spins are aligned in the

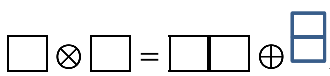


positive *z*-direction. You can swap the order of any pair of electrons, that is, and not change the symmetry of the eigenstates. As for the totally antisymmetric states we may try vertical tableaux like

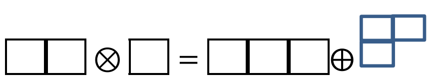
but these are illegal as the numbers must increase as we go down.

We now define a mixed symmetry tableau like  . We may visualize such a tableau as either a singlet box attached to a symmetric tableau, or as a single box attached to an antisymmetric tableau:  . Let's work a few examples.

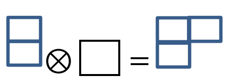
We've seen $\frac{1}{2} \otimes \frac{1}{2} = 1 \oplus 0$. This translates to 2×2=3+1. In terms of Young's tableaux we have

 ; $2 \times 2 = 3 + 1$ (triplet + singlet).

What about $1 \otimes \frac{1}{2} = \frac{3}{2} \oplus \frac{1}{2}$? We have

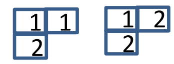
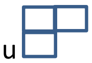
 ; $3 \times 2 = 4 + 2$ (quartet+doublet).

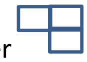
Lastly consider $0 \otimes \frac{1}{2} = \frac{1}{2}$. We have

 ; $1 \times 2 = 2$ (doublet).

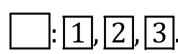
If the spin function for three electrons is symmetric in two of the indices, neither of them can be antisymmetric with respect to a third index, *e.g.*, $(|+\rangle_1|-\rangle_2 + |-\rangle_1|+\rangle_2)|-\rangle_3$ is symmetric with $1 \leftrightarrow 2$

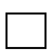
but is neither symmetric wrt $1 \leftrightarrow 3$ or $2 \leftrightarrow 3$. If the number cannot decrease in the horizontal direction

and must increase in the vertical direction, the only possibilities are . The tableau 

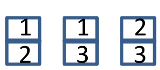
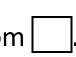
therefore corresponds to a doublet. Notice that in this scheme we do not consider .

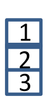
When we were studying the relationship of two spin-$\frac{1}{2}$ objects between the coupled and uncoupled representations with ladder operators or Young's tableaux above, we were really studying the group SU(2). The two particles could have been the z-direction spin labels of two electrons, or they could have been the labels up and down for the *u* and *d* quarks. We now extend our Young's tableaux considerations to three primitive spin-$\frac{1}{2}$ objects—think SU(3). The labels 1, 2 and 3 may stand for the magnetic quantum numbers of p-orbitals in atomic physics or charge states of the pion π^+, π^0, π^-, or the u, d and s quarks in the SU(3) classification of elementary particles. A box can assume three possibilities:



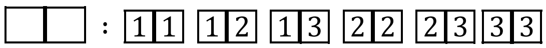
The dimensionality of  is 3. Here is some the SU(3) Young's tableaux algebra if you will:

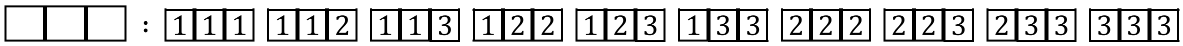
<u>Antisymmetry</u>:

 : : dimensionality 3. Note that in SU(3) we use 3^* to distinguish  from .

 :  : dimensionality 1 (This state is totally antisymmetrical.)

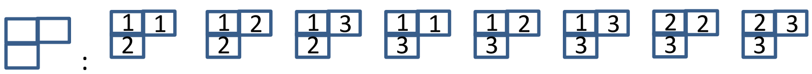
<u>Symmetry</u>:

 : dimensionality 6.

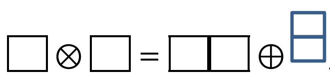
 :

dimensionality 10.

<u>Mixed symmetry</u>:

 : dimensionality 8.

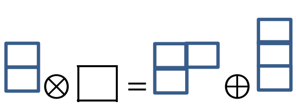
Sakurai develops a general tableaux for representations of SU(3) along with a formula for the dimensionality of a given tableaux which is not necessary to these notes. Let's practice with a few problems.

 ; $3 \times 3 = 6 + 3$ (6 symmetric, 3 antisymmetric).

Using the methods we had before Young's tableaux we would have written

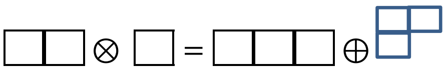
$$1 \otimes 1 = 2 \oplus 1 \oplus 0,$$

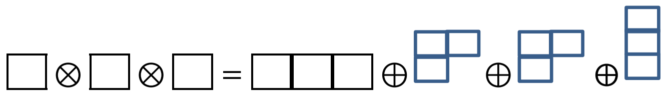
or, changing this to ordinary products and sums: $3 \times 3 = 5 + 3 + 1$. The advantage of Young's tableaux is the sum 9 is broken into symmetric, mixed symmetric and antisymmetric bins—there were no mixed symmetry states in this example. Now let's check out:

 ; $3 \times 3 = 8 + 1$ (8 mixed symmetry, 1 totally antisymmetric). Our old methods would have

$$1 \otimes 1 = 2 \oplus 1 \oplus 0,$$

It's the same result as before $(3 \times 3 = 5 + 3 + 1)$ but Young's tableaux tells us more: 8 mixed symmetry states this time with 1 totally antisymmetric state. Let's do two more examples.

 ; $6 \times 3 = 10 + 8$ (6 symmetric, 8 mixed symmetry). Lastly,



$$3 \times 3 \times 3 = 10 + 8 + 8 + 1.$$

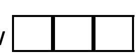
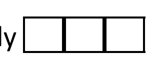
We can refine $10 = 7 + 3, 8 = 5 + 3, 8 = 5 + 3$ in terms of angular momentum states. We have,

$j = 3$ (dimension 7) once (totally symmetric),

$j = 2$ (dimension 5) twice (both mixed symmetry),

$j = 1$ (dimension 3) three times (one totally symmetric, two mixed symmetry),

$j = 0$ (dimension 1) (totally antisymmetric state).

The $j = 0$ state is unique, corresponding to the fact that the only product of vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ invariant under rotation is $\mathbf{a} \cdot (\boldsymbol{b} \times \mathbf{c})$, which is necessarily antisymmetric. (You should read Sakurai's fascinating material connecting the symbol  to an antisymmetric state $j = 1$ behaving like the ordinary vector cross product $\mathbf{a} \times \mathbf{b} = \mathbf{c}$. He also discusses how  breaks down, noting that it must contain $j = 3$, but that this is not the only symmetric state; $\mathbf{a}(\mathbf{b} \cdot \mathbf{c}) + \mathbf{b}(\mathbf{c} \cdot \mathbf{a}) + \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$ is totally symmetric, having the transformation properties of $j = 1$. Evidently  contains both the $j = 3$ (seven states) and $j = 1$ (three states).

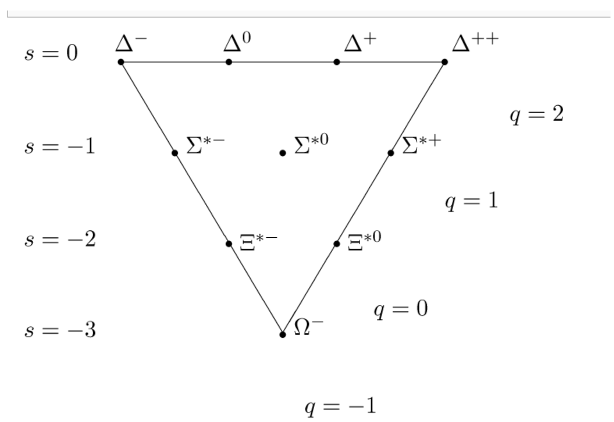
Optional diversion on color. This is the last and optional material from Sakurai before getting to the optional Wigner-Eckert theorem and the material on weight and root diagrams and Dynkin diagrams. The SU(3) decuplet corresponds to $\square\square\square = (\mathbf{10})$

$$\Delta^{++}, \Delta^{+}, \Delta^{0}, \Delta^{-1} \quad ddd, udd, uud, uuu \;\; I = \frac{3}{2}$$

$$\Sigma^{+}, \Sigma^{0}, \Sigma^{-1} \quad dds, uds, uus \;\; I = 1$$

$$\Xi^{0}, \Xi^{-} \quad dss\, uss \;\; I = \frac{1}{2}$$

$$\Omega^{-} \quad sss \quad I = 0.$$



All of these ten states are known to be spin $\frac{3}{2}$ objects. It's safe to assume (Sakurai's words) that the space part is in a relative S-state for low lying states of three quarks. We expect total symmetry in the spin degree of freedom. For example, the $j = \frac{3}{2}\, m = \frac{3}{2}$ state of Δ can be visualized to have quark spins all aligned, BUT the quarks are spin $\frac{1}{2}$ particles subject to Fermi-Dirac statistics. Yet with

Quark label (now called flavor): symmetric

Spin: symmetric

Space: symmetric

for the $j = \frac{3}{2}$ decuplet. The total symmetry is even—the historical "statistics paradox"! The way out of this was to postulate that there is an additional degree of freedom called color (red, blue or yellow) and postulate that the observed hadrons (the strongly interacting particles including the $J = \frac{3}{2}^{+}$ states considered here are color singlets

$$\frac{1}{\sqrt{6}}(|RBY\rangle - |BRY\rangle + |BYR\rangle - |YBR\rangle + |YRB - |RYB).$$

This is in complete analogy with the unique  totally antisymmetric combination in color space. The statistics problem is now solved because

$$P_{ij} = P_{ij}^{(flavor)} P_{ij}^{(spin)} P_{ij}^{(space)} P_{ij}^{(color)}, \boldsymbol{or}$$

$$(-) = (+)(+)(+)(-).$$

This "cheap" solution in fact comes with measurable consequences such as the decay rate of the π^0 and the cross section for electron-positron annihilation into hadrons. End of optional Sakurai based SU(3) particle physics material.

(Optional) **Wigner-Eckert theorem on selection rules**. In quantum mechanics we typically have to evaluate quantities of the nature $\langle\psi|A|\phi\rangle$ where A is some operator and ψ and ϕ are eigenstates of an unperturbed Hamiltonian H_o which is invariant under some symmetry; for now consider invariance under SO(3).

<u>Scalar case</u>. Let $\mathbf{X}$ be an angular momentum operator (spin or angular momentum) with components X_1, X_2, X_3 (we may see $\mathbf{X}$ as a vector thusly) and let S be a scalar operator. Then

$[\mathbf{X}, S] = 0$, or equivalently wrt SO(3) rotations $U(R)$, $U(R)SU(R)^{-1} = S$. (Weight and length, for example, are invariant under rotations.) From the commutator we may extract selection rules that Unless $j = j'$ and $m = m'$, $\langle j'm'|S|jm\rangle = 0$. Moreover, this matrix element is independent of m, so that $\langle j'm'|S|jm\rangle = N_j \delta_{jj'} \delta_{mm'}$ (no summation). Let's show these conclusion more explicitly.

The condition $j = j'$ is a consequence of the vanishing of the commutator $[S, \mathbf{X}^2]$. Let me explain. Consider

$$\langle j'm'|S\mathbf{X}^2|jm\rangle = j(j+1)\langle j'm'|S|jm\rangle.$$

But since S is just a scalar (a number), $S\mathbf{X}^2 = \mathbf{X}^2 S$. So

$$\langle j'm'|\mathbf{X}^2 S|jm\rangle = j'(j'+1)\langle j'm'|S|jm\rangle.$$

It must also be that $\langle j'm'|S\mathbf{X}^2|jm\rangle = \langle j'm'|\mathbf{X}^2 S|jm\rangle$ so

$$j(j+1)\langle j'm'|S|jm\rangle = j'(j'+1)\langle j'm'|S|jm\rangle.$$

Equivalently

$$(j - j')(j + j' + 1)\langle j'm'|S|jm\rangle = 0.$$

Certainly $(j + j' + 1)$ is positive. So either $j = j'$ or $\langle j'm'|S|jm\rangle$ is zero. Now for showing that $m = m'$. This follows from $[S, X_3] = 0$. That is, $(SX_3 - X_3 S)\psi = S(X_3 - X_3)\psi = 0$. Lastly, $\langle j'm'|S|jm\rangle$ depends only on j and not on m. This is a consequence of the vanishing of $[S, X_\pm]\psi$. Consider

$$\langle jm|S|jm\rangle = \left\langle j'm'\middle|SX_\pm\middle|jm\right\rangle = \left\langle j'm'\middle|X_\pm S\middle|jm\right\rangle.$$

Then

$$\sqrt{(j-m+1)(j+m)}\langle jm|S|jm\rangle = \langle jm|SX_+|jm-1\rangle$$

and, recalling that $(X_+)^\dagger = X_-$,

$$\langle jm|SX_+|jm-1\rangle = \langle jm|X_+S|jm-1\rangle = \sqrt{(j-m+1)(j+m)}\langle jm-1|S|jm-1\rangle.$$

This is the Wigner-Eckert theorem for scalars. QED.

Vector case. Before getting to the theorem, let's build up our understanding of vector operators. Suppose there is an observable $\mathbf{V}$ that is a vector. That is, its three components V_x, V_y, V_z in an orthonormal frame satisfy

$$[J_x, V_x] = 0, \qquad [J_x, V_y] = i\hbar V_z, \qquad [J_x, V_y] = -i\hbar V_y \quad \text{(A)}$$

as well as those obtained by cyclic permutations of the indexes x, y and z. See complement $\mathrm{B_{VI}}$ of "Quantum Mechanics, Volume Two", Claude Cohen-Tannoudji, Bernard Diu and Franck Laloë. (The material of the vector case is from complement $\mathrm{D_X}$.) To give an idea of what these above commutators mean, let's study some examples of vector operators.

The angular momentum $\mathbf{J}$ is itself a vector. If we let $\mathbf{V}$ be $\mathbf{J}$ the above commutators become the usual angular momentum commutation relations, *e.g.*, $[J_x, J_y] = i\hbar J_z$. On the other hand, if we let $\mathbf{J}$ be $\mathbf{L}$ (orbital angular momentum) and $\mathbf{V}$ be $\mathbf{R}$ a radius vector, then the relations (A) become

$$[L_x, X] = [YP_z - zP_y, X] = 0,$$

$$[L_x, Y] = [YP_z - zP_y, Y] = [-ZP_y, Y] = i\hbar Z,$$

$$[L_x, Z] = [YP_z - zP_y, Z] = [YP_z, Z] = -i\hbar Y.$$

For a particle of spin $\mathbf{S}$, $\mathbf{J}$ is given by $\mathbf{J} = \mathbf{L} + \mathbf{S}$. In this case the operators $\mathbf{L}, \mathbf{S}, \mathbf{R}, \mathbf{P}$ are vectors. If we take into account the fact that all the spin operators (which act only on the spin state space) commute with the orbital operators (which act only in the $\mathbf{R} = \{X, Y, Z\}$ state space), the proof of the commutator relations (A) follows from our first two examples. On the other hand, operators of the type $\mathbf{L}^2$, $(\mathbf{L} \cdot \mathbf{S})$, etc., are not vectors, but scalars. Other vector operators could, however, be constructed from these scalars, *e.g.*, $\mathbf{R} \times \mathbf{X}$ and $(\mathbf{L} \cdot \mathbf{S})\mathbf{P}$, etc.

Consider the system (1) + (2) formed by the union of two systems: (1), of state space $\mathcal{E}_1$. And (2), of spates space $\mathcal{E}_2$. If $\mathbf{V}(1)$ is an operator which acts only in $\mathcal{E}_1$, and if this vector (that is, satisfies (A) with angular momentum vector $\mathbf{J}_1$, of the first system), then the extension of $\mathbf{V}(1)$ into state space $\mathcal{E}_1 \otimes \mathcal{E}_2$ is also a vector. (This tensor product stuff is what we have been doing.) For example, for a two-electron system, the operators $\mathbf{L}_1, \mathbf{R}_1, \mathbf{S}_2$, etc. are vectors.

Wigner-Eckert theorem for vectors. We introduce the operators V_+, V_-, J_+, J_- defined by

$$V_\pm = V_x \pm iV_y$$

$$J_\pm = J_x \pm iJ_y.$$

Using relations (A) we can show that:

$$[J_x, V_\pm] = \mp \hbar V_{z,}$$

Let's check this.

$$[J_x, V_\pm] = [J_x, V_x \pm iV_y] = [J_x, V_x] \pm i[J_x, V_y] = 0 \pm i \cdot i\hbar V_z = \mp V_z.$$

Similarly,

$$[J_y, V_\pm] = -i\hbar V_z,$$

$$[J_z, V_\pm] = \pm \hbar V_\pm.$$

Once we have verified the above commutation relations, we can show

$$[J_+, V_+] = 0,$$

$$[J_+, V_-] = 2\hbar V_z,$$

$$[J_-, V_+] = -2\hbar V_z,$$

$$[J_-, V_-] = 0.$$

This kind of algebra, by the way, is the kind of algebra that underlies root and weight diagrams and Dynkin diagrams that underlie the Standard Model and extensions beyond it using larger Lie groups, *e.g.*, SU(5) GUT and Lissi E8, which are defunct, but very instructive.

Let's show that $\langle jm|V_\pm|j'm'\rangle = 0$ whenever $m \neq m'$. To show this it suffices to note that V_z and J_z commute (you can prove this to yourself). Therefore the matrix elements of V_z between two vectors $|jm\rangle$ corresponding to two different eigenvalues of $m\hbar$ of J_z are zero.

For the matrix elements $\langle jm|V_\pm|j'm'\rangle$ of $V_\pm$ we shall show that they are different from zero only if $m - m' = \pm 1$. The equation $[J_z, V_\pm] = \pm\hbar V_\pm$ indicates that:

$$J_z V_\pm = V_\pm J_z \pm \hbar V_\pm.$$

Applying both sides of this relation to the ket $|j'm'\rangle$ we obtain

$$J_z(V_\pm)|j'm'\rangle = V_\pm J_z|j'm'\rangle \pm \hbar V_\pm|j'm'\rangle = (m' \pm 1)\hbar V_\pm|j'm'\rangle.$$

Evidently $V_\pm|j'm'\rangle$ is an eigenvector of J_z with eigenvalues $(m' \pm 1)\hbar$. Since two eigenvectors of a the Hermitian operator J_z associated with different eigenvalues are orthogonal, it follows that the scalar product $\langle jm|V_{\pm}|j'm'\rangle$ is zero if $m \neq m' \pm 1$. Summing up, the selection rules obtained for the matrix elements of $\mathbf{V}$ are as follows:

$$V_z \Rightarrow \Delta m = m - m' = 0,$$

$$V_+ = \Delta m = m - m' = +1,$$

$$V_- = \Delta m = m - m' = -1.$$

QED. The matrix associated with V_z is diagonal and those associated with $V_{\pm}$ have matrix elements only just above and just below the principal diagonal.

A similar procedure can be adopted to find the selection rules for a tensor operator, $\mathbf{T}$. However the procedure rapidly becomes cumbersome and a more powerful method is required, namely the Wigner-Eckert theorem (see H F Jones, "Groups, Representation and Physics," 2nd. Ed. An irreducible tensor operator T_m^l is an operator that transforms under rotations according to

$$U(R)T_m^l U(R)^{-1} = D_{mm'}^j(R)|jm\rangle.$$

(Tensor operators may have more indices .) Stating the result, the Wigner-Eckert theorem states that

$$\langle jm|T_m^l|j'm'\rangle = C(Jj'j; Mm'm)\langle j||T^j||j'\rangle.$$

The $C(Jj'j; Mm'm)$ are our Clebsch-Gordan coefficients.

Root and Weight diagrams and Dynkin diagrams. Before the theory, let's begin with an example that will make the theory transparent. In quantum physics we seek to discover the largest possible set of commuting hermitian generators because we want to diagonalize as much as possible. Such a largest possible subset is called the **Cartan subalgebra**.

Example 1—Consider the eight Gell-Mann matrices from Part I, the defining representation of SU(3).

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \qquad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \qquad \lambda_8 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

These matrices are hermitian ($\lambda_i = \lambda_i^\dagger$) and traceless. For the Gell-Mann matrices, the Cartan subalgebra is easy to pick out: λ_3 and λ_8 are the only diagonal matrices, and hence they commute: $[\lambda_3, \lambda_8] = 0$. Therefore we would say in quantum physics that λ_3, λ_8 have simultaneously measurable eigenvalues. The dimension of the Cartan subalgebra (being 2) is called the rank of SU(3).

Let's consider the commutator algebra $[\lambda_i, \lambda_j]$. Find the commutator of the first two matrices:

$$[\lambda_1, \lambda_2] = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 2i\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 2i\lambda_{3.}$$

You can check for yourself that $[\lambda_2, \lambda_1] = -2i\lambda_3$. In general, $[\lambda_i, \lambda_j] = \frac{1}{2} f_{ijk}\lambda_k$, where the repeated index implies summation from $k = 1,2, \ldots, 8$, and the f_{ijk} are the structure constants. For the Gell-Mann matrices, (dividing the matrices by $1/\sqrt{2i}$) we have

$$f_{123} = 1, \quad f_{458} = f_{678} = \frac{\sqrt{3}}{2}, \quad f_{177} = f_{165} = f_{246} = f_{257} = f_{345} = f_{376} = 1/2.$$

Any f_{ijk} involving an i, j, k permutation not on this list is zero. Even permutations, *e.g.*, $123 \rightarrow 312 \rightarrow 213$ have the same value. The odd permutations $(123 \rightarrow 132)$ have the opposite sign. So $f_{123} = f_{312} = f_{231} = -1$ and $f_{132} = f_{321} = f_{213} = -1$. The symbol f_{ijk} is said to be totally antisymmetric in i, j, k.

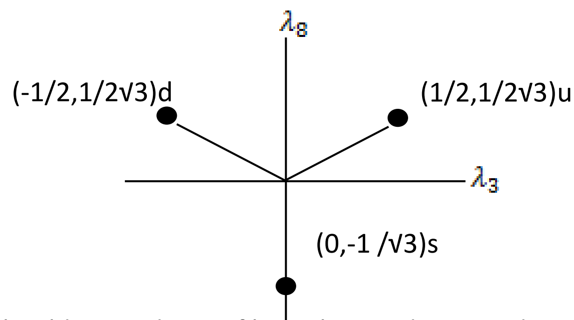
Example 2—Consider

$$[\lambda_4, \lambda_8] = 2if_{45a}\lambda_a = 2i[f_{453}\lambda_3 + f_{458}\lambda_8] = 2i\left[\frac{1}{2}\lambda_3 + \frac{\sqrt{3}}{2}\lambda_8\right] = 2i\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

We could have gone the long way to obtain this result by computing the matrices directly in $\lambda_4\lambda_8 - \lambda_8\lambda_4$. Georgi's text uses $T_a = \frac{1}{2}\lambda_a$. Let's use this "T" relabeling. The eigenvectors of T_3 and T_8 are

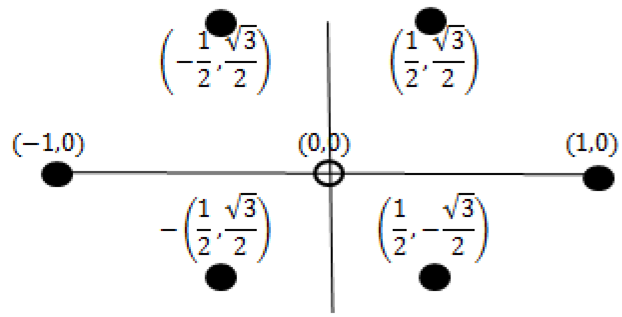
$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

with eigenvalues, respectively, of $\frac{1}{2}, -\frac{1}{2}$ and 0 for T_3 and $\frac{\sqrt{3}}{6}, \frac{\sqrt{3}}{6}, -\frac{\sqrt{3}}{3}$. They plot like:

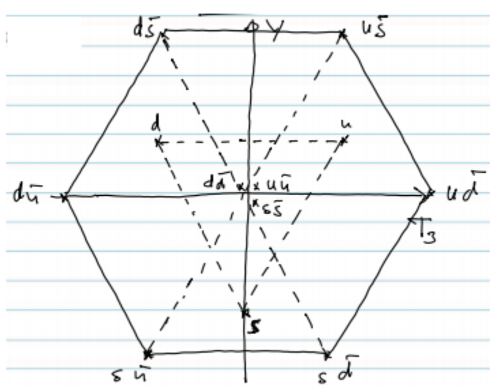


This graph (**weight diagram**) should remind you of how the up, down, and strange quarks fill a graph of hypercharge versus the third component of Isospin T_3. The Gell-Mann matrices are a particular representation of $SU(3)$, the special unitary matrices with determinant +1. Finally note that the Gell-Mann matrices are a generalization of the Pauli spin matrices in the sense that the first three Gell-Mann matrices contain the Pauli matrices acting on a subspace $\lambda_a = \begin{pmatrix} \sigma_a & 0 \\ 0 & 0 \end{pmatrix}$, $a = 1,2,3$.

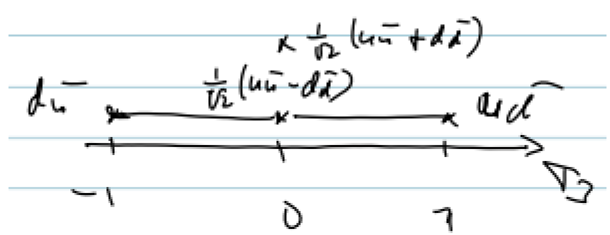
The root diagram is easy to construct from the weight diagram. Let $\mathbf{A} = \left(-\frac{1}{2}\hat{\imath}, \frac{1}{2\sqrt{3}}\hat{\jmath}\right), \mathbf{B} = \left(\frac{1}{2}\hat{\imath}, \frac{1}{2\sqrt{3}}\hat{\jmath}\right), \mathbf{C} = \left(0\hat{\imath}, -\frac{1}{\sqrt{3}}\hat{\jmath}\right)$. Start taking differences of these vectors. $\mathbf{A} - \mathbf{B} = (-1\hat{\imath}, 0\hat{\jmath})$, $\mathbf{A} - \mathbf{C} = \left(-\frac{1}{2}\hat{\imath}, \frac{\sqrt{3}}{2}\hat{\jmath}\right)$, $\mathbf{B} - \mathbf{C} = \left(\frac{1}{2}\hat{\imath}, \frac{\sqrt{3}}{2}\hat{\jmath}\right)$, $\mathbf{B} - \mathbf{A} = (1\hat{\imath}, 0\hat{\jmath})$, $\mathbf{C} - \mathbf{A} = \left(\frac{1}{2}\hat{\imath}, -\frac{\sqrt{3}}{2}\hat{\jmath}\right)$, $\mathbf{C} - \mathbf{B} = \left(-\frac{1}{2}\hat{\imath}, -\frac{\sqrt{3}}{2}\hat{\jmath}\right)$. We can also get the vector $\mathbf{0}$ three different ways. The following figure is our **root diagram**.



This is reminiscent of $\square \otimes \square = \square\square \oplus \begin{smallmatrix}\square \\ \square\end{smallmatrix}$; $3 \times 3 = 6 + 3$ (6 symmetric, 3 antisymmetric). In fact, the two previous figures lead to a graphical representation of SU(3). In terms of the superposition of the previous two drawings (three quarks for the weight diagram superposed with the meson nonet):



The dimension of SU(3) is the rank of the Cartan subalgebra, namely 2, plus the number of root vectors extending from the origin, namely 6. That is $\dim \mathrm{SU}(3) = 2 + 6 = 3^2 - 1$. Note quickly that when we only had the SU(2) up and down quarks we could have built a graphical representation of SU(2) thusly:



See: http://theory.gsi.de/~friman/e-part-script/EPP_11.pdf .

We may get the same diagramatic weigth and root results in terms of ladder operators (raising and lower operators). This is a more official way. Consider

$$[T_3, E_\alpha] = \alpha_i E_\alpha$$

where T_3 belongs to the Cartan subalgebra and E_α is a raising or lowering operator. Let's make this explicit. Using Georgi's matrix notation for the Gell-Mann matrices we have:

$$T_1 = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_2 = \frac{1}{2}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T_4 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$T_5 = \frac{1}{2}\begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad T_6 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad T_7 = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad T_8 = \frac{1}{2\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Now we build our raising and lowering operators E_α and $E_\alpha^\dagger$ for $SU(3)$ as follows. Let:

$$T_1 + iT_2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad T_1 - iT_2 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$T_4 + iT_5 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad T_4 - iT_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

$$T_6 + iT_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad T_6 - iT_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Now let: $E_{\pm 1,0} = \frac{1}{\sqrt{2}}(T_1 \pm T_2)$, and $E_{\pm\frac{1}{2},\pm\frac{\sqrt{3}}{2}} = \frac{1}{\sqrt{2}}(T_4 \pm T_5), etc.$ Let's see what this mess we've made does for us. (You'll soon see that these operators move us around the figure above which is reminiscent of the pseudoscalar meson nonet.)

$$[T_3, T_1 + iT_2] = [T_3, E_\alpha] = \left[T_3, E_{+1,0}\right] = \frac{1}{2}\begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = T_1 + iT_2 = 1 \cdot E_{+1,0}.$$

So the eigenvalue is $\alpha = 1$. Consider $[T_8, T_1 + iT_2] = \left[T_8, E_{+1,0}\right] = 0 \cdot (T_1 + iT_2) = 0 \cdot E_{+1,0}; \;\therefore\; \alpha = 0$ for this case. So we get the point $(1,0)$ of our meson nonet. Let's try a few more.

$$[T_3, T_1 - iT_2] = -(T_1 - iT_2), \therefore \;\; \alpha = -1, \qquad [T_8, T_1 - iT_2] = 0 \cdot (T_1 - iT_2), \therefore \;\alpha = 0.$$

Now we have the point $(-1,0)$ of our nonet. Let's proceed to:

$$[T_3, T_4 + iT_5] = (T_4 + iT_5), \therefore \;\; \alpha = \frac{1}{2}, \qquad [T_8, T_4 + iT_5] = \frac{\sqrt{3}}{2}(T_4 + iT_5), \therefore \;\; \alpha = \frac{\sqrt{3}}{2}, \ldots,$$

and finally,

$$[T_3, T_3] = [T_3, T_8] = [T_8, T_8] = 0,$$

the three points at the origin. Doing all the cases gives us our complete nonet root diagram. Let me just say that the Killing form is used to transform the Cartan subalgebra into an orthonormal basis. We'll get to that in a moment, but first let me show you yet one other way to build our nonet via the adjoint representation of SU(3).

The adjoint representation of $\mathbf{SU(3)}$. The structure constants themselves generate a representation of the algebra called the ***adjoint representation***. That is, we may define a set of 8×8 matrices $[T_a]_{bc} = -if_{abc} \;\; \ni \;\; [T_a, T_b] = if_{abc}$ just as the Gell-Mann matrices do.

Example 3—Let's see.

$$T_3 = \begin{pmatrix} 0 & 1 = f_{312} & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 = f_{321} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} = f_{345} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} = f_{354} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} = f_{367} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} = f_{876} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$T_8 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{\sqrt{3}}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{\sqrt{3}}{2} = f_{367} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{\sqrt{3}}{2} = f_{876} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The matrices in the adjoin representation are 7×7 since f_{8ab} with either a of $b = 8$ give zero. The eigenvectors for T_3 with eigenvalues as subscripts are:

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}_0, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ -\sqrt{2}i \\ -\sqrt{2} \\ 0 \\ 0 \\ 0 \end{bmatrix}_{-\frac{1}{2}}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ -\sqrt{2}i \\ \sqrt{2} \\ 0 \\ 0 \\ 0 \end{bmatrix}_{\frac{1}{2}}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \sqrt{2}i \\ -\sqrt{2} \\ 0 \end{bmatrix}_{-\frac{3}{2}}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \sqrt{2}i \\ \sqrt{2} \\ 0 \end{bmatrix}_{\frac{3}{2}}, \quad \begin{bmatrix} -\sqrt{2}i \\ -\sqrt{2} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{-1}, \quad \begin{bmatrix} -\sqrt{2}i \\ \sqrt{2} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}_1.$$

For T_8 we have:

$$\begin{bmatrix}0\\0\\1\\0\\0\\0\\0\\0\end{bmatrix}_0, \quad \begin{bmatrix}0\\0\\0\\-\sqrt{2}i\\-\sqrt{2}\\0\\0\\0\end{bmatrix}_{-\frac{\sqrt{3}}{2}}, \quad \begin{bmatrix}0\\0\\0\\-\sqrt{2}i\\\sqrt{2}\\0\\0\\0\end{bmatrix}_{\frac{\sqrt{3}}{2}}, \quad \begin{bmatrix}0\\0\\0\\0\\0\\\sqrt{2}i\\-\sqrt{2}\\0\end{bmatrix}_{\frac{\sqrt{3}}{2}}, \quad \begin{bmatrix}0\\0\\0\\0\\0\\-\sqrt{2}i\\-\sqrt{2}\\0\end{bmatrix}_{\frac{\sqrt{3}}{2}}, \quad \begin{bmatrix}-1\\0\\0\\0\\0\\0\\0\\0\end{bmatrix}_0, \begin{bmatrix}0\\1\\0\\0\\0\\0\\0\\0\end{bmatrix}_0.$$

The first three eigenvectors for T_3 and T_8 are simultaneous. So too are the next two because they differ only by the factor -1 and this doesn't matter (when you're squaring things). The last two eigenvalues of T_8 do not appear (at first glance) to be simultaneous eigenvectors with the last two eigenvectors of T_3. Notice, however, that the last eigenvectors of T_8 have the same eigenvalue, namely 0. As you learn in quantum physics, we may proceed to break this degeneracy by a change of basis. Consider the subspace of

$$T_3: \begin{pmatrix}0 & 1\\-1 & 0\end{pmatrix}$$

with normalized eigenvectors

$$\begin{pmatrix}\frac{1}{\sqrt{2}}i\\\frac{1}{\sqrt{2}}\end{pmatrix}, \quad \begin{pmatrix}-\frac{1}{\sqrt{2}}i\\\frac{1}{\sqrt{2}}\end{pmatrix}.$$

We may form a unitary matrix U such that $UU^\dagger = \mathbb{I}_{2\times2}$. That is:

$$U = \frac{1}{\sqrt{2}}\begin{pmatrix}i & -i\\1 & 1\end{pmatrix}, \qquad U^\dagger = \frac{1}{\sqrt{2}}\begin{pmatrix}-i & 1\\i & 1\end{pmatrix}, \quad UU^\dagger = \begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}.$$

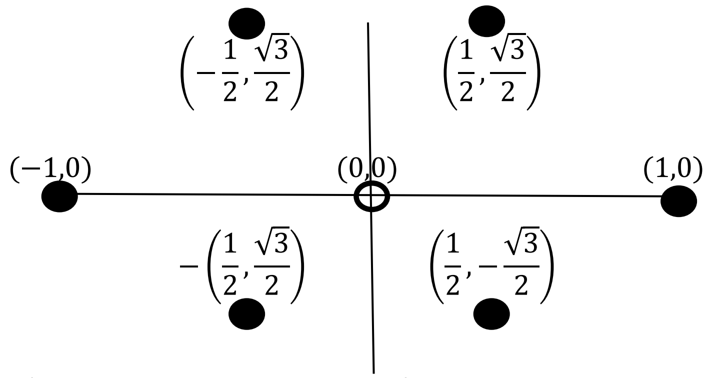
Then

$$U^\dagger\begin{pmatrix}0 & 1\\-1 & 0\end{pmatrix}U = \begin{pmatrix}-i & 0\\0 & i\end{pmatrix} = \text{Diagonal matrix.}$$

The eigenvectors with eigenvalues as subscripts are

$$\begin{pmatrix} i \\ 0 \end{pmatrix}_{-1}, \qquad \begin{pmatrix} 0 \\ i \end{pmatrix}_{1}.$$

In this basis, the subspace of T_3 has simultaneous eigenvectors with T_8. If we plot these eigenvalues of T_3 against T_8 as coordinate pairs we again get our nonet:



The eigenvalues in the adjoint representation are named ***roots***.

The Cartan subalgebra for SU(3) was of rank 2. The roots and weights can be plotted on plane. The roots and weights of other, larger Lie algebras require three, four, or many more dimensions to plot. The first grand unified theory (1974) (a unified theory of our known forces excluding gravitation) was based on SU(5), the Standard Model being a subgroup of this larger Lie group. People tried to find a larger group (SU(5)) than contained the Standard Model groups. This embedding led to new vertices in the Feynman rules corresponding to a more general, unified Lagrangian. The group theory tells you about what transforms to what under interaction, but does not say very much about the actual functional form of the potentials. In this vein in 2007, Lissi published a Theory of Everything, a theory which includes gravitation, based on the exceptional Lie group E8. Many tools have been developed to the study of the four infinite Lie algebras as well as the five exceptional Lie algebras. Among these tools are Dynkin diagrams introduced in 1947 by Eugene Dynkin. A Dynkin diagram records the configuration of a Lie algebra's simple roots. I have given you what you need to read the readable work of Wangberg and Dray in "Visualizing Lie Subalgebras Using Root and Weight Diagrams," A Wangberg and T Dray, Department of Mathematics, Oregon State University, Corvallis, Oregon 97331, 15 March 2008. You may find this brief but explicit article at: http://www.math.oregonstate.edu/~tevian/JOMA/joma_paper_softlinks.pdf . This material is brought to life by the authors at http://www.math.oregonstate.edu/~tevian/JOMA/joma_paper_three.html . You can see the plotting being done by JAVA script. These authors show you how to build Dynkin diagrams, and how to use them to help you study subalgebras. I have pitched my copy of Georgi in the trash, something I should have done years ago. H F Jones is now far more readable. This concludes the "Lie" program tying differential equations (ordinary and partial, linear and nonlinear) with algebra and global and local topology with physics.

Part V.I Fractional calculus. Differentiation and integration can be generalized beyond integrating and differentiating in integer steps, *e.g.*, 1st derivative, 2nd derivative, and so on. I recommend "The Fractional Calculus, Theory and Applications of Differentiation and Integration to Arbitrary Order," K. B. Oldham and J. Spanier, Dover, 1974. It is far more general and applied and full of centuries of history than a website I found on this subject. Nevertheless, the notes I made from the website provide a good beginning to Oldham and Spanier.

Consider $f(x) = e^{ax}$. If you take n derivatives you get: $\frac{d^n}{dx^n} f(x) = a^n e^{ax}$. (We'll soon see that if n is a negative integer, this leads to integrations.) So why don't we replace n with ν where ν is not necessarily a positive integer? Then, $\frac{d^\nu}{dx^\nu} = a^\nu e^{\nu x}$. In fact any function that is expressible as a linear combination of exponential functions can be differentiated likewise:

$$\frac{d^\nu}{dx^\nu}\cos x = \frac{d^\nu}{dx^\nu}\left[\frac{e^{ix} - e^{-ix}}{2}\right] = \frac{(i)^\nu e^{ix} + (-i)^\nu e^{-ix}}{2}.$$

Since $(\pm i)^n = e^{\pm \frac{n\pi}{2}}$,

$$\frac{d^\nu}{dx^\nu}\cos x = \cos\left(x + \frac{\nu\pi}{2}\right).$$

The differential operator shifted the phase of the cosine function (ditto the sine function).

Now consider that a very large class of functions have Fourier representations:

$$f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(\alpha) e^{-i\alpha x} d\alpha,\ where\ \ g(\alpha) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x) e^{i\alpha x} dx.$$

Here f and g are Fourier transforms of each other. Then differentiating $f(x)$ by d^ν/dx^ν results in

$$\frac{d^\nu}{dx^\nu}f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-} g(\alpha)(-i\alpha)^\nu e^{-i\alpha x}d\alpha.$$

Evidently the Fourier transform of the generalized n^{th} of $f(x)$ is $(-i\alpha)^\nu g(\alpha)$. This approach to the fractional calculus has pathologies. With this approach, how do we determine the half-derivative of $f(x) = x$? There is no Fourier representation for this open-ended function. We could chose a finite interval, but which one? Let's turn back to the definition of differentiation.

$$\frac{d}{dx}f(x) = \lim_{\epsilon\to 0}\frac{f(x) - f(x-\epsilon)}{\epsilon}.$$

Repeat once to get:

$$\frac{d^2}{dx^2} = \lim_{\epsilon\to 0}\frac{\frac{f(x)-f(x-\epsilon)}{\epsilon} - \left(\frac{f(x-\epsilon)-f(x-2\epsilon)}{\epsilon}\right)}{\epsilon} = \lim_{\epsilon\to 0}\frac{f(x) - 2f(x-\epsilon) + f(x-2\epsilon)}{\epsilon^2}.$$

Repeat again to get:

$$\frac{d^3}{dx^3} = \lim_{\epsilon\to 0}\frac{\frac{f(x) - f(x-\epsilon) - 2f(x-\epsilon) + 2f(x-2\epsilon) + f(x-2\epsilon) - f(x-3\epsilon)}{\epsilon^2}}{\epsilon}$$

$$= \lim_{\epsilon\to 0}\frac{f(x) - 3f(x-\epsilon) + 3f(x-2\epsilon) - f(x-3\epsilon)}{\epsilon^3}.$$

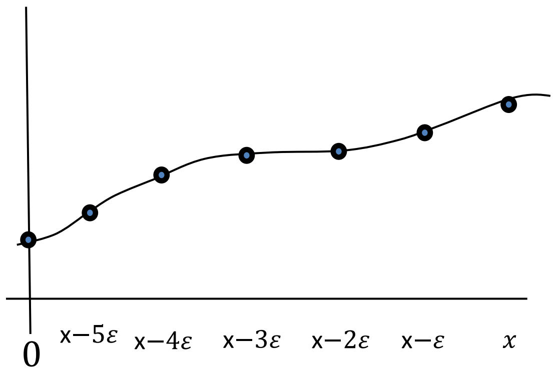
Repeating n times yields

$$\frac{d^n}{dx^n}f(x) = \lim_{\epsilon\to 0}\frac{1}{\epsilon^n}\sum_{j=0}^{n}(-1)^j\binom{j}{n}f(x-j\epsilon). \tag{1}$$

Indeed if $f(x) = x^4$ (dropping all powers of ϵ^2 or higher):

$$\frac{d^2}{dx^2}x^4 = \lim_{\epsilon\to 0}[x^4 - 2(x-\epsilon)^4 + (x-2\epsilon)^4] = \lim_{\epsilon\to 0}[12x^2 - 24x\epsilon + 14\epsilon^2] = 12x^2.$$

We generalize (1) for non-integer orders, generalizing the binomial coefficients and determining the generalization of the upper summation limit. Consider a smooth function $f(x)$ as shown below:



Note that $x = k\epsilon = 5\epsilon$. Define the shift operator $\sigma_\epsilon f(x) = f(x-\epsilon)$. Consider an operator D

$$D^n[f(x)] = \lim_{\epsilon\to 0}\left(\frac{1-\sigma_\epsilon}{\epsilon}\right)^n f(x).$$

What is $D^1 f(x)$?

$$D^1 f(x) = \lim_{\epsilon\to 0}\left(\frac{1-\sigma_\epsilon}{\epsilon}\right) f(x) = \lim_{\epsilon\to 0}\left(\frac{f(x)-f(x-\epsilon)}{\epsilon}\right) = \frac{d}{dx}f(x).$$

What is $D^{-1}f(x)$?

$$D^{-1}f(x) = \lim_{\epsilon\to 0}\left(\frac{1-\sigma_\epsilon}{\epsilon}\right)^{-1} f(x) = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\cdot\frac{1}{1-\sigma_\epsilon}f(x).$$

Recall that $\frac{1}{1-z} = 1 + z + z^2 + z^3 + \cdots$. Then (with $\epsilon = \frac{x}{N-1}$)

$$D^{-1}f(x) = \lim_{\epsilon\to 0}\epsilon(1+\sigma_\epsilon+\sigma_\epsilon^2+\cdots)f(x) = \lim_{\epsilon\to 0}\epsilon\big(f(x)+f(x-\epsilon)+f(x-2\epsilon)+\cdots+f(0)\big)$$

$$= \lim_{\epsilon\to 0}\sum_{k=0}^{N-1} f(x-k\epsilon)\epsilon = -\int_x^0 f(x)dx = \int_0^x f(x)dx.$$

“Differentiating” to -1 order is integrating once!

Regarding equation (1), if n is a positive integer, all of the binomial coefficients after the first $n+1$ are zero ($C(n;j) = 0 \ \forall \ j > n$. For negative or fractional values of n the binomial coefficients are non-terminating. We must sum over the specified range. The upper limit should be $\frac{x-x_0}{\epsilon}$ yielding:

(1a)

$$\frac{d^n}{dx^n} f(x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon^n} \sum_{j=0}^{\left|\frac{(x-x)}{\epsilon}\right|} (-1)^j \binom{j}{n} f(x - j\epsilon).$$

What do we do about $\binom{j}{n} = \frac{n!}{j!(n-j)!}$ when n is not an integer? Enter the gamma function. For integers

$$\Gamma(n) = (n-1)!.$$

It is a consequence of the general gamma function for complex-valued arguments, which is

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt.$$

Most special functions of applied mathematics arise as solutions to differential equations. Not so the gamma function—at least not yet (Hölder's theorem). From the following integral

$$\int_{-1}^{+1} (1-x^2)^n dx = \frac{(2n+1)!\,(n!)^2}{2^{2n+1}}, \qquad \text{we get,} \qquad n! = \sqrt{\frac{(2n+1)!}{2^{2n+1}} \int_{-1}^{+1} (1-x^2)^n dx}\,.$$

Thanks to the $(2n+1)$ argument, we have generalized the definition of the factorial to negative and/or positive half-integers, e.g., $\left(-\frac{1}{2}\right)! = \sqrt{\pi}$. But we have actually done better than this. Now that the factorial of all half-integers is defined, we can compute the factorial of quarter-integers, sixteenth integers,…,so that, using the binary representation of real numbers, and using the identity the $(x+1)! = (x+1)x!$, we actually have a well-defined factorial function for any real number, otherwise known as the gamma function: $\Gamma(n) = (n-1)! \ \forall \ n > 0$. The recurrence relationship is $\Gamma(x+1) =$

$x\Gamma(x)$. The reflection relation is $\Gamma(x)\Gamma(1-x) = p/\sin(px)$. One could write volumes on the gamma function. Now we can generalize (1a) to

$$\frac{d^\nu}{dx^\nu} f(x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon^\nu} \sum_{j=0}^{\left|\frac{(x-x)}{\epsilon}\right|} (-1)^j \frac{\Gamma(\nu+1)}{\mathrm{j}!\,\Gamma(\nu+1-\mathrm{j})} f(x - j\epsilon). \tag{2}$$

Ordinary differentiation is local, integration is not local, nor is our generalized definition of differentiation/integration (2) (See Oldham and Spanier). If n is integer-valued, the binomial coefficients for all j vanish past $j = n$, and in the limit as ϵ goes to zero the n values of $f(x - k\epsilon)$ with nonzero coefficients converge on x, keeping the generalized operator local. In general, however, the binomial expansion has infinitely many non-zero coefficients, so the result depends on the value of x all the way down to x_0. (Look up Huygens' Principle and the sharp propagation of light in three dimensions.) Choosing $x_0 = 0$, equation (2) becomes

$$\frac{d^\nu}{dx^\nu} f(x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon^\nu} \sum_{j=0}^{\left|\frac{x}{\epsilon}\right|} (-1)^j \frac{\Gamma(\nu+1)}{\mathrm{j}!\,\Gamma(\nu+1-\mathrm{j})} f(x - j\epsilon). \tag{3}$$

Let's go for a test drive with $f(x) = x$ and $n = 1/2$.

$$\frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}} x$$

$$= \lim \frac{1}{\sqrt{\epsilon}} \left[1 \cdot \frac{\Gamma(\frac{3}{2})}{0!\,\Gamma(\frac{3}{2})} x - 1 \cdot \frac{\Gamma\left(\frac{3}{2}\right)}{1!\,\Gamma\left(\frac{1}{2}\right)} (x - \epsilon) + 1 \cdot \frac{\Gamma\left(\frac{3}{2}\right)}{2!\,\Gamma\left(-\frac{1}{2}\right)} (x - 2\epsilon) - \begin{pmatrix} \frac{1}{2} \\ \left[\frac{x}{\epsilon}\right] - 2 \end{pmatrix} (2\epsilon) - \begin{pmatrix} \frac{1}{2} \\ \left[\frac{x}{\epsilon}\right] - 1 \end{pmatrix} (\epsilon) \right] x$$

$$\epsilon \to \infty$$

$$= \lim_{\epsilon \to \infty} \frac{1}{\sqrt{\pi\epsilon}} \left[1\sigma_\epsilon - \frac{1}{2}\sigma_\epsilon^2 - \frac{1}{8}\sigma_\epsilon^3 - \cdots \right] x$$

where the factorials are in terms of the gamma function. Recall that we used $\left(\frac{1-\sigma_\epsilon}{\epsilon}\right)^n$. With $n = \frac{1}{2}$

$$\left(\frac{1-\sigma_\epsilon}{\epsilon}\right)^{1/2} = \frac{1}{\sqrt{\epsilon}}\sqrt{1-\sigma_\epsilon} = \frac{1}{\sqrt{\epsilon}}\left(1 - \frac{1}{2}\sigma_\epsilon - \frac{1}{8}\sigma_\epsilon - \cdots\right).$$

Then really have:

$$\frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}x = \lim_{\epsilon\to\infty}\frac{1}{\sqrt{\pi\epsilon}}\sqrt{1-\sigma_\epsilon}x = \lim_{\epsilon\to\infty}\sqrt{\frac{x^2-\sigma_\epsilon x^2}{\pi\epsilon}} = \lim_{\epsilon\to\infty}\sqrt{\frac{x^2-(x-2\epsilon)^2}{\pi\epsilon}} = \lim_{\epsilon\to\infty}\sqrt{\frac{x^2-x^2+4x\epsilon-4\epsilon^2}{\pi\epsilon}}$$

$$= \lim_{\epsilon\to\infty}\sqrt{\frac{4x\epsilon-4\epsilon^2}{\pi\epsilon}} = 2\sqrt{x/\pi}.$$

That is,

$$\frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}x = 2\sqrt{\frac{x}{\pi}}. \tag{4}$$

If we play around a bit with regular calculus we observe the pattern $\frac{d^n}{dx^n}x^m = \frac{m!}{(m-n)!}x^{m-n}$. In our generalized framework we have

$$\frac{d^\nu}{dx^\nu}x^m = \frac{m!}{\Gamma(m+1-\nu\,)}x^{m-\nu}. \tag{5}$$

Then, paradoxically, using equation (5), the half derivative of a constant c is:

$$\frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}c = c\cdot\frac{0!}{\Gamma\left(\frac{1}{2}\right)}\sqrt{x} = \frac{c}{\sqrt{\pi x}}.$$

We have to be careful and think about local versus nonlocal behavior. Equation (5) gives no hint on non-locality, as opposed to equation (3) which depends on what x is doing locally. This issue is cleared up by considering the half derivative of the exponential function e^{ax}. Using equation (5) we get

$$\frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}e^x = \frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}\left(1+x+\frac{1}{2!}x^2+\cdots\right) = \frac{1}{\sqrt{\pi x}}\left(1+2x+\frac{4}{3}x^3+\frac{16}{105}x^4+\cdots\right).$$

Wouldn't we have expected this derivative of the exponential function to equal itself? A plot of the above series versus e^x shows that the two graphs converge asymptotically at large x. However near the origin, the series blows up whereas e^x passes smoothly through $x = 0$. Consider the integrals

$$\int_{x_0}^{x} u^3 du = \frac{x^4}{4} - \frac{x_0^4}{4}, \qquad \int_{x_0}^{x} e^u du = e^x - e^{x_0}.$$

The left hand integral shows that when we say x^3 is the derivative of $\frac{x^4}{4}$ we are implicitly assuming $x_0 = 0$. This is consistent with our derivation of (3). HOWEVER, when we say that e^x is the derivative of e^x we are saying that $x = -\infty$ so that$e^{x_0} = 0$. Thus for each of the two above integrals/derivatives we have tacitly assumed different ranges. The fix is to go back to equation (2) and replace $x_0 = 0$ with $x_0 = -\infty$. Now we're cooking:

$$\frac{d^\nu}{dx^\nu}f(x) = \lim_{\epsilon\to 0}\frac{1}{\epsilon^\nu}\sum_{j=0}^{\infty}(-1)^j\frac{\Gamma(\nu+1)}{\mathrm{j}!\,\Gamma(\nu+1-\mathrm{j})}f(x-j\epsilon).$$

One point of this section is to out the glaring omission of calculus with other than integer order differentiation and integration, it can be real-value-ordered, even complex-value-ordered. There are also strong applications of fractional calculus to the problems of transport and diffusion processes such as neutron migration and flow through porous materials. Talking about diffusion processes…

Part V.II Stochastic differential equations (SDEs), random walks and the Itô calculus. My goal here is to prove a theorem concerning quadratic variation. Why? For practical ends I assure you. How does one integrate a function in which some of the terms are stochastic? In difference form consider

$$\Delta x = a(x,t)\Delta t + b(x,t)\Delta W,$$

where $\Delta W = N(0,1)$, the normal distribution with mean 0 and standard deviation 1. We start at $x = 0, t = 0$ and head to t_{final}ending up at some unknown value of x. What would be our "expectation" of x? (What is the mean of x? What is the standard deviation of x? The physicists answered this question employing Monte Carlo methods requiring computers. With a computer we throw a dart, so-to-speak and pick a number between 0 and 1 according to $N(0,1)$ (or some other distribution). We advance x and repeat, generating a path until we arrive at t_{final}. The we start the process from scratch, repeating it many times over so that we obtain a numerical estimate of the mean and standard deviation of x. The mathematicians did things nice and proper in a measure-theoretic framework. Is there a middle ground? Yes. The material presented below. Along the way we will see applications to diffusion and high finance.

On quadratic variation (this first portion excerpted from the Feynman lectures). Let D_N be the net distance traveled in N steps (each with time t_0 and speed v_0) of a drunken sailor in one-dimension. Then, (with expectations, or means in brackets),

$$D_{t_0} = v_0 t_0 = l_0, \qquad D_{t_0}^2 = l_0^2, \qquad \langle D_{t_0}^2 \rangle = l_0^2.$$

If the sailor steps to the right, then we have

$$D_{Nt_0} = D_{(N-1)t_0} + l_0, \qquad D_{Nt_0}^2 = D_{(N-1)t_0}^2 + 2D_{(N-1)t_0} + l_0^2.$$

If the sailor steps to the left (with equal probability) then we have

$$D_{Nt_0} = D_{(N-1)t_0} - l_0, \qquad D_{Nt_0}^2 = D_{(N-1)t_0}^2 - 2D_{(N-1)t_0} + l_0^2.$$

The sum $D_{Nt_0}^2(\text{right}) + D_{Nt_0}^2(\text{left})$ is then

$$2D_{Nt_0}^2 = 2D_{(N-1)t_0}^2 = 2l) + 2l_0^2.$$

(The twos can be removed.) Since $\langle D_{t_0}^2 \rangle = l_0^2$,

$$\langle D_{2t_0}^2 \rangle = \langle D_{t_0}^2 \rangle + l_0^2 = 2l_0^2,$$

$$\langle D_{3t_0}^2 \rangle = \langle D_{2t_0}^2 \rangle + l_0^2 = 3l_0^2, \ldots,$$

$$\langle D_{Nt_0}^2 \rangle = Nl_0^2.$$

This expresses the variance. The root-mean-square (rms) (or standard deviation) distance is thus

$$D_{\text{rms}} = \sqrt{N}l_0 = \sqrt{N}v_0 t_0 = \sqrt{Nt_0}v_0\sqrt{t_0}.$$

Note that Nt_0 is the total walking time, T. If we set $t_0 = 1s, v_0 = 1\ m/s$ then $l_0 = 1m$. Thus

$$D_{\text{rms}} = \sqrt{T}[m].$$

The standard deviation of the distance (in meters) after N steps is proportional to the square root of the time—the key idea to quadratic variation. Note—The universal character of Brownian motion is simply the dynamic counterpart (evolution in time) of the universal nature of its static counterpart, the normal or Gaussian distribution. Both arise from the same source, the central limit theorem. This theorem states that when we average large numbers of independent and comparable objects, we obtain the normal distribution. That is, what we observe is the result of a very large number of individually very small influences. Note that $2D_{Nt_0} = 2D_{(N-1)t_0} + l_0 - l_0$. Thus $\langle D_{Nt_0} \rangle = \langle D_{(N-1)t_0} \rangle \Rightarrow \text{(induction)}$ $\langle D_{t_0} \rangle = l_0$. On average the sailor goes nowhere, his standard deviation goes like $\sqrt{t}[m]$.

Necessary review of probability theory. Let's start by reviewing **moments**. If $f(x)$ is a probability density function (density or distribution) the first moment about the origin (the center of mass) is

$$\mu_1 = E[x] = \int_{-\infty}^{\infty} f(x)dx.$$

The first moment about μ_1 is $E[x - E(x)] = 0$. The second moment (the VARIANCE) about the mean is $E[(x-\mu)^2] = E[x^2 - 2x\mu + \mu^2] = E[x^2 - 2\mu^2 + \mu^2] = E[x^2 - \mu^2] = E[x^2 - (E[x])^2] = \sigma^2$. The r^{th} moment about the mean is defined by

$$\mu_r = E[(x-\mu)^r] = \int_{-\infty}^{\infty} (x-\mu)^r f(x)dx.$$

Naturally, the r^{th} the moment about the origin is $E[x^r]$. <u>In general</u> $\mathrm{Var}[x^r] = E[x^{2r} - (E[x^r])^2]$.

The Moment Generating Function (MGF). Let X be a random variable. The MGF of X is $M_X(t) = E[e^{tX}]$ at all values of t for which the expected value exists. In integral form,

$$M_X(t) = \int_{-\infty}^{\infty} e^{tX} f(x)dx.$$

Example 1—Let our density be

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, \text{otherwise} \end{cases}$$

$$M_X(t) = \int_{-\infty}^{\infty} e^{tX} e^{-x} dx = \frac{1}{1-t}.$$

Theorem (no proof)—The r^{th} derivative of the MGF obeys $M_X^{(r)}(0) = \int_{-\infty}^{\infty} x^r f(x)dr = E[x^r]$. ***<u>NOTE</u>***: the expectation $E[x^r]$ are taken with $t = 0$ in $M_X(t)$.

Example2—For example 1: $M_X^{(r)}(t) = \frac{r!}{(1-t)^{-r-1}}$, so $M_X^{(r)}(0) = r! = E[x^r]$. This gives our mean and variance. The mean is $\mu = E[x^1] = M_X^{(1)}(0) = 1! = 1$. The variance is $\text{Var}(x) = E(x^2) - \mu^2 = M_X^{(2)}(0) - 1 = 2! - 1 = 1$.

Now we will denote the MGF by:

$$M_X(t) \equiv E[e^{tx}] = \int_{-\infty}^{\infty} e^{tx} dF(x).$$

Note that if X and Y are independent random variables, $M_{X+Y} = M_X(t) \cdot M_Y(t)$. Think convolution.

$$M_X(t) \equiv E[e^{tx}] = 1 + tE[x] + \cdots + \frac{t^k}{k!} E[x^k] + e(t),$$

where $e(t)$ denotes the round off error. As before $M_X^{(k)}(t) = E[x^k]$ for the k^{th} derivative. Consider $N(0,1) = f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, (**A**) a normal density with mean 0 and standard deviation 1. Let $x' = \mu + \sigma x$, then we have $N(\mu, \sigma)$ being defined by

$$f(x') = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\frac{1}{2}(x'-\mu)^2}{\sigma^2}}.$$

(***<u>Drop the primes</u>***.) Getting back to the MGF for $N(\mu = 0, \sigma = 1)$,

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} N(0,1) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{1}{2}x^2} dx.$$

Completing the square: $tx - \frac{1}{2}x^2 = -\frac{1}{2}(x-t)^2 + \frac{1}{2}t^2$, so (**B**)

$$M_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{1}{2}x^2} dx = \frac{e^{\frac{1}{2}t^2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} dx = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}t^2}.$$

The last integral is so because with $x' = x - t,\ dx' = dx$, we get $\int_{-\infty}^{\infty} e^{-\frac{1}{2}x'^2} dx' = 1$. Thus for $N(0,1),\ M_X(t) = \int_{-\infty}^{\infty} e^{tx} N(0,1)dx = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}t^2} \equiv E[e^{tx}]$. Combining (**A**) with (**B**) we get:

$$E\left[e^{t(\mu+\sigma x)}\right] = e^{\mu t} E[e^{\sigma t x}] = e^{\mu t} e^{\frac{1}{2}\sigma t^2}. \ \text{ thus, }\ M_{x=\mu+\sigma t}(t) = E\left[e^{t(\mu+\sigma t)}\right] = e^{\mu t + \frac{1}{2}\sigma^2 t^2},$$

but now for the normal density $N(\mu, \sigma)$. We will see this result when we deal with a simple stochastic differential equation (SDE) for Brownian motion (or stocks). Let $\mu = 0$.

$$M_{\sigma t}(t) = e^{\frac{1}{2}\sigma^2 t^2} = 1 + \frac{1}{2}\sigma^2 t^2 + \frac{1}{2!}\left(\frac{1}{2}\sigma^2 t^2\right)^2 + O(t^2) = 1 + \frac{1}{2}\sigma^2 t^2 + \frac{3}{4!}\sigma^4 t^4 + O(t^2)$$
$$= \sum_{r=0}^{\infty} \frac{M_X(0)}{r!} t^r.$$

So the variance of x with this density is $E[x^2] = \mathrm{Var}(x) = \sigma^2$, and $E[x^4] = 3\sigma^4$, …, and so on by ***reading off*** the successive terms (***the statistical moments about the mean***) from the MGF series. Thus the ***var***iance of X^2 is $E[X^2] = \mathrm{Var}(X^2) = E[X^4] - (E[X^2])^2 = 2\sigma^4$. Let's put all of this crap together for once and all.

A ***Brownian motion*** (drunken sailor). For each $t > 0$ the random variable $W(t) = W(t) - W(0)$ is the (drunken) increment in time $[0, t]$ where the increments are Gaussian increments with mean zero, variance t, and with density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\frac{1}{2}(x-\mu)^2}{\sigma^2}} \longmapsto f(t, x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{\frac{1}{2}x^2}{t}} = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}.$$

That is, $\Delta W = W(t_0 + t) - W(t_0) \backsim N\left(0, \sqrt{t}\right)$. (What does this? Drunken sailors walking in one or more dimensions, neutrons diffusing in a star, and approximately stock prices. Physicists extract the mean (first moment about the origin) and variance (second moment about the mean) by Monte Carlo.

By definition we have set $E[W(t)] \equiv 0$. By definition we have set $\sigma_{\mathrm{W(t)}} \equiv \sqrt{t}$. Then plugging into the definition of variance of $W(t)$, we have $\mathrm{Var}(W(t)) \equiv E[W^2(t) -$ ~~$(E[W(t)])^2$~~$] = E[W^2(t)] = \sigma^2_{\mathrm{W(t)}} = t$. Plugging into the definition of variance for $(W^2(t))$, we have $\mathrm{Var}(W^2(t)) \equiv E[W^4(t) - (E[W^2(t)])^2] = E[W^4(t) - t^2] = E[3t^2 - t^2] = 2t^2$. I got the $3t^2$ term by translating the result for $E[X^2] = \mathrm{Var}(X^2) = E[X^4] - (E[X^2])^2 = 2\sigma^4$ above. (Remember that in a Brownian process (Feynman's drunken sailor $\sigma^2 = t$. So $\sigma^4 = t^2$.)

Thus for $0 < t \ll 1,\ t \gg t^2,$ and $\mathrm{Var}(W^2(t)) = 2t^2 \ll \mathrm{Var}(W(t)) = E[W^2(t)] = t$. With a fine enough partition of P_n of a time interval $[0, t]$ we see what is called the <u>quadratic variation of Brownian motion</u>.

$$\sum_{i=1}^{n} [\Delta W(t_i)]^2 = \sum_{i=1}^{n} \Delta t_i = t.$$

Theorem—$\sum_{i=1}^{n} [\Delta W(t_i)]^2 = \sum_{i=1}^{n} \Delta t_i = t$ in probability $\ni \max|t_i - t_{i-1}| \to 0\ \forall P_n$. (Proof see above—screw measure theory in the absence of context).

This result is used in a key theorem of the Itô calculus and in measure theoretic integration theory. Let's get to Itô's lemma. (A Weiner process/Brownian motion are the same thing.)

Itô's lemma. Given a Weiner process $\Delta W = N(0,1)\sqrt{\Delta t}$, let $\Delta x = a(x,t)\Delta t + b(x,t)\Delta W$. Then $E[\Delta x] = a\Delta t$ since $b(x,t)E[\Delta W] = 0$, and $\mathrm{Var}(\Delta x) = b^2\Delta t$. Thus $\sigma = \sqrt{\mathrm{Var}(\Delta x)} = b\sqrt{t}$. Say $G = G(x, y)$ with no stochastic attribute for now. Then

$$\Delta G = \frac{\partial G}{\partial x}\Delta x + \frac{\partial G}{\partial y}\Delta y + \frac{1}{2}\cdot\frac{\partial^2 G}{\partial x^2}\Delta x^2 + \frac{1}{2}\cdot\frac{\partial^2 G}{\partial x \partial y}\Delta x\Delta y + \frac{1}{2}\cdot\frac{\partial^2 G}{\partial y^2}\Delta y^2 + \text{Higher Order (H.O. terms)}.$$

For small $\Delta x, \Delta y,\ \Delta G \approx \frac{\partial G}{\partial x}dx + \frac{\partial G}{\partial y}dy$. Now suppose $\Delta x = (x,t)\Delta t + b(x,t)\mathrm{N}(0,1)\sqrt{\Delta t}$. Then

$$\Delta x^2 = a^2(x,t)\Delta t^2 + abN(0,1)\Delta t^{\frac{3}{2}} + b^2N^2(0,1)\Delta t.$$

With $\Delta t^2 \ll \Delta t^{\frac{3}{2}} \ll \Delta t,\ \Delta x^2 \approx b^2N^2(0,1)\Delta t$. Thus $E[\Delta x^2] = \mathrm{Var}(\Delta x) = b^2\Delta t$. Also $\mathrm{Var}(\Delta x^2) = E\left[b^4N^4(0,1)\Delta t^4 - \left(E\left[b^2N^2(0,1)\right)\Delta t\right)^2\right] = 2b^4\Delta t^2$ (we've just derived this above, but now the time interval Δt is included). As a result of this (and $\Delta t \gg \Delta t^2$ for small t) we may treat Δx^2 as non-stochastic with mean $b^2 dt$ as $\Delta t \to 0$. (We dropped the variance term of Δx thanks to the quadratic variation of the random walk we are working with.) With $G = G(x,t)$, taking limits as $\Delta x \to dx,\ \Delta t \to dt$ we get (KEEPING ONLY FIRST POWERS OF DIFFERENTIALS, e.g., dx, dt but dropping $dx^2, dy^2, dxdy$)

$$dG = \frac{\partial G}{\partial x}dx + \frac{\partial G}{\partial t}dt + \frac{1}{2}\cdot\frac{\partial^2 G}{\partial x^2}dx^2 = \frac{\partial G}{\partial x}dx + \frac{\partial G}{\partial t}dt + \frac{1}{2}\cdot\frac{\partial^2 G}{\partial x^2}b^2 dt.$$

Cool result. With $dx = adt + bdW$, we get:

$$dG = \frac{\partial G}{\partial x}(adt + bdW) + \frac{\partial G}{\partial t}dt + \frac{1}{2}\cdot\frac{\partial^2 G}{\partial x^2}b^2 dt = \left(\frac{\partial G}{\partial x}a + \frac{\partial G}{\partial t} + \frac{1}{2}\cdot\frac{\partial^2 G}{\partial x^2}b^2\right)dt + \frac{\partial G}{\partial x}bdW.$$

So what? Itô's lemma allows us to deduce the moments of an arbitrary function $G(x,t)$ where x is following a stochastic Brownian motion! This is powerful. Stock prices can be approximated by Brownian motion. Derivative instruments are financial instruments defined as functions of time t and the stochastic (Brownian process) price x. The drift rate of $G(x,t)$ is

$$\frac{\partial G}{\partial x}a + \frac{\partial G}{\partial t} + \frac{1}{2}\cdot\frac{\partial^2 G}{\partial x^2}b^2$$

(and looking at the coefficient for dW) the variance rate is

$$\frac{\partial^2 G}{\partial x^2}b^2.$$

This is one of the theoretical bases for pricing derivatives (sensible derivatives based on some stochastic process which doesn't have to be Brownian, a Langevin process say. Derivatives based on BS are dangerous.) Other ways to extract moments are Monte Carlo simulation (you must have good random number generators), binary trees, and trinary trees (equivalent to finite difference methods for PDEs). There is a beautiful mapping of the heat equation to the Black-Scholes equation for call and put options.

Example 3—A variation of an Orhstein-Uhlenbeck (father of the H-bomb) mean reverting process. Let

$$dx(t) = -\beta x(t)dt + \sqrt{2\beta\sigma^2}dW, \qquad x(0) = x_o, \qquad \beta, \sigma = \text{constants.}$$

Find the mean, variance and probability density function of the process analytically—no computers!

$$\langle dx \rangle = -\beta\langle x \rangle dt + 0$$

since $\langle dW \rangle = 0$.

$$\int_{\langle x_0 \rangle}^{\langle x \rangle} \frac{d\langle x \rangle}{\langle x \rangle} = -\beta \int_{t=0}^{t} dt.$$

Thus, $\langle x(t) \rangle = \langle x_0 \rangle e^{-\beta t}$, the mean is $E\langle x \rangle = E\langle x_0 \rangle e^{-\beta t}$. The variance is $\mathrm{Var}[x] = \langle x^2 \rangle - \langle x \rangle^2$. Recall Itô's lemma. With $dx(t) = a(x(t), t)dt + b(x(t), t)dW$,

$$dF[x(t), t] = \left(\frac{\partial}{\partial t} + a\frac{\partial}{\partial x} + \frac{\frac{1}{2}b^2\partial^2 F}{\partial x^2} \right) dt + b\frac{\partial F}{\partial x}dW.$$

Comparing the equation for $dx(t)$ at the top with the equation just above this line, we see that

$$a = -\beta x, \qquad b = \sqrt{2\beta\sigma^2}.$$

If $F[x(t), t] = x^2$ say, (with $DF = dx^2$) then

$$dF = \left(0 + 2ax + \frac{1}{2}b^2 \cdot 2\right)dt + b \cdot 2xdW = (2ax + b^2)dt + 2bxdW.$$

Plugging our results for a and b into the equation above leads to

$$dF = dx^2 = (-2\beta x^2 + 2\beta\sigma^2)dt + 2\sqrt{2\beta\sigma^2}xdW.$$

So,

$$\langle dx^2\rangle = -2\beta\langle x^2\rangle dt + 2\beta\sigma^2 dt + 2\sqrt{2\beta\sigma^2}\langle \cancel{x^2 dW}\rangle$$

$$\langle dx^2\rangle = -2\beta\langle x^2\rangle dt + 2\beta\sigma^2 dt.$$

Then

$$\frac{d\langle x^2\rangle}{dt} + 2\beta\langle x^2\rangle = 2\beta\sigma^2.$$

Thus with integrating factor: $\mu = e^{\int_0^t 2\beta t} = e^{2\beta t}$ we have

$$e^{2\beta t}\frac{d\langle x^2\rangle}{dt} + e^{2\beta t}\langle x^2\rangle = 2\beta\sigma^2 e^{2\beta t}.$$

So $\frac{d}{dt}\left[\langle x^2\rangle e^{2\beta t}\right] = 2\beta\sigma^2 e^{2\beta t}$. Integrating we get:

$$\int d\langle x^2 e^{2\beta t}\rangle = 2\beta\sigma^2 \int_0^t e^{2\beta t} = 2\beta\sigma^2\left[\frac{1}{2\beta}\left(e^{2\beta t} - 1\right)\right].$$

So $\langle x^2\rangle e^{2\beta t} - c = \sigma^2\left(e^{2\beta t} - 1\right)$, where c is a constant. Now the initial condition $\langle x^2(0)\rangle = \langle x_0^2\rangle = \sigma^2 \cdot 0 + ce^{-2\beta t.}$ So, $E[x^2] = \langle x^2\rangle = \sigma^2\left(1 - e^{-2\beta t}\right) + \langle x_0^2\rangle e^{-2\beta t}$. Therefore $\text{Var}[x] = E[x^2] - (E[x])^2 = \sigma^2 + \left(\sigma^2 + \langle x_0\rangle^2 e^{-2\beta t}\right) + \langle x_0\rangle^2 e^{-2\beta t} = \sigma^2 + (\langle x_0^2\rangle - \langle x_0\rangle^2 - \sigma^2)e^{-2\beta t}$. So $\text{Var}[x] = \text{Var}[x_0 - \sigma^2]e^{-2\beta t} + \sigma^2$. We seek the density associated with the corresponding Fokker-Planck equation.

Some (optional) Background on Fokker-Planck equation. ("A modern course in statistical physics," L. E. Reichl, University of Texas Press, 1980.) (Graduate school level statistical physics.) The Master equation is

$$\frac{\partial P_1}{\partial t} = \int \{W(y_1, y_2)P_1(y_1(t) - W_2(y_2, y_1)P_1(y, t)\}\, dy,$$

where P_1 is the density for a Brownian particle and W accounts for the conditional, or transmitted density per unit time. This process is ***stationary*** if $P_1(y, t) = P(y_1)$ (no time dependence). If the transitions W are small we get the Fokker-Planck equation

$$\frac{\partial P_1}{\partial t} = -\frac{\partial}{\partial t}\left[\alpha_1(y)P_1(y, t) + \frac{1}{2} \cdot \frac{\partial^2}{\partial y^2}[\alpha_2(y)P_2(y, t)]\right],$$

Where α_n is the n^{th} order jump moment. If particle collisions are restricted to being binary, we get the restricted Boltzmann (Vlasov) equation:

$$\frac{\partial f}{\partial t} + \vec{\mathbf{u}} \cdot \nabla f + \frac{\vec{\mathbf{F}}}{m} \cdot \nabla_u f = \int d^3\alpha_1 \int d\Omega |\vec{\mathbf{u}} - \vec{\mathbf{u}}_\mathbf{1}| E[\Omega](f f_1' - f f_1).$$

Detailed balance means $f f_1 = f f_1'$. This leads to the H-theorem (which leads to the approach to equilibrium). A Hamiltonian may supply the transition matrix, *e.g.*, BBGKY. END.

The Fokker-Planck equation (FPE) governs the transport of the probability density for a given stochastic process. For a general diffusion process, $dx_i = D_i dt + E dW_i$.

This process maps to the FPE as

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x_i}(D_i P) + \frac{1}{2}\frac{\partial^2}{\partial x_i \partial x_j}\left(P E^2 \delta_{ij}\right).$$

For our Orhstein-Uhlenbeck process $dx = -\beta x dt + \sqrt{2\beta\sigma^2}dW$, we have $D = -\beta x$, $E = \sqrt{2\beta\sigma^2}$, and

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}(-\beta x P) + \frac{1}{2}\cdot\frac{\partial^2}{\partial x^2}(2\beta\sigma^2 P).$$

We shall solve this PDE by the Fourier transform method.

$$\mathfrak{F}\left[\frac{\partial P}{\partial t}\right] = \mathfrak{F}\left[\beta\frac{\partial}{\partial x}(xP)\right] + \frac{1}{2}\cdot\mathfrak{F}\left[\frac{\partial^2}{\partial x^2}(2\beta\sigma^2 P)\right]$$

$$\frac{\partial \hat{P}}{\partial t} = \beta\mathfrak{F}\left[\frac{\partial f}{\partial x}\right] + \beta\sigma^2(-ik)\hat{P},$$

where $f = \frac{\partial}{\partial x}(xP)$. Note that $\mathfrak{F}\left[\frac{\partial f}{\partial x}\right] = (-ik)\hat{F} = (-ik)[xP(x)] = (-ik)\left[\frac{\partial \hat{P}}{\partial k}\right]$, where $\mathfrak{F}[g(x)] = \int_{-\infty}^{\infty} g(x)e^{ikx}dx = \hat{g}(x)$. So our equation becomes

$$\frac{\partial \hat{P}}{\partial t} + \beta k\frac{\partial \hat{P}}{\partial k} = -\beta\sigma^2\hat{P},$$

a first-order ODE. The method of characteristics is appropriate. Let $\hat{P}$ be constant along the characteristics

$$u(k, t, \hat{P}) = a = \text{constant},$$

$$v(k, t, \hat{P}) = b = \text{constant}.$$

Then,

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial k}\cdot\frac{\partial k}{\partial t} + \frac{\partial u}{\partial \hat{P}}\cdot\frac{\partial \hat{P}}{\partial t} = 0,$$

$$\frac{dv}{dt} = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial k} \cdot \frac{\partial k}{\partial t} + \frac{\partial v}{\partial \hat{P}} \cdot \frac{\partial \hat{P}}{\partial t} = 0.$$

Along the characteristics we require the reparameterization

$$\frac{d\hat{P}}{dt} = \beta\sigma^2 k^2 \hat{P}, \qquad \frac{dk}{dt} = \beta k.$$

This implies that $dt = \frac{dk}{\beta k}$, or

$$dt = \frac{d\hat{P}}{\beta\sigma^2 k^2 \hat{P}}$$

along the characteristic. Check: Plug $\frac{d\hat{P}}{dt}$, $\frac{dk}{dt}$, dt into the equations for $\frac{du}{dt}$ and $\frac{dv}{dt}$.

$$\frac{\partial u}{\partial t} + \beta k \frac{\partial u}{\partial k} + \beta\sigma^2 \cdot \frac{\partial u}{\partial \hat{P}} \hat{P} = 0,$$

$$\frac{\partial v}{\partial t} + \beta k \frac{\partial v}{\partial k} + \beta\sigma^2 \cdot \frac{\partial v}{\partial \hat{P}} \hat{P} = 0.$$

Now if $u = \hat{P} = v = \hat{P} =$ constant along the characteristic, we see that our pair of equations above becomes

$$\frac{\partial \hat{P}}{\partial t} + \beta k \cdot \frac{\partial \hat{P}}{\partial k} - \beta\sigma^2 k^2 \hat{P} = 0,$$

Recovering the original equation. Therefore along these characteristics a general solution exits as $f(u, v) = 0$. We get

$$dt = \frac{dk}{\beta k}.$$

Thus $k = e^{\beta t}$. Also

$$\frac{dk}{\beta k} = -\frac{d\hat{P}}{\beta\sigma^2 k^2 \hat{P}}.$$

Thus $\sigma^2 k dk = -d\hat{P}/\hat{P}$. So $\ln(\hat{P}) = -\frac{1}{2}\sigma^2 k^2$ and $\hat{P} = \hat{P}_0 e^{-\frac{\sigma^2 k^2}{2}}$. WLOG let $\hat{P}_0 = 1$. Let $a = b = 1$, then since $u = u(k, t, \hat{P})$ and $v = v(k, t, \hat{P})$, set

$$u(t, k, \hat{P}) = e^{\beta t} e^{-\beta t} = k e^{-\beta t} = 1,$$

$$v(t, k, \hat{P}) = e^{-\frac{\sigma^2 k^2}{2}} e^{\frac{\sigma^2 k^2}{2}} = \hat{P} e^{\frac{\sigma^2 k^2}{2}} = 1.$$

The general solution can be written as $v = g(u)$ where $g(u)$ is an arbitrary function. We have $\hat{P}(k, t) = e^{-\frac{\sigma^2 k^2}{2}} g(k e^{-\beta t})$. With initial condition (I.C.) $\hat{P}(x, 0|x_0, t_0) = \delta(x - x_0)$, we get

$$\hat{P}(k, t) = e^{-\frac{\sigma^2 k^2}{2}} g(k) = e^{ikx_0}.$$

This implies that $g(k) = e^{\frac{\sigma^2 k^2}{2} + ikx_0}$. Finally, $\hat{P}(k, t|k_0, t_0) = e^{-\frac{\sigma^2 k^2}{2}(1 - e^{-2\beta\Delta t}) + ikx_0 e^{-\beta\Delta t}}$. Invert this

$$P(x, t|x_0, t_0) = \int_{-\infty}^{\infty} e^{-\frac{\sigma^2 x}{2}(1 - e^{-2\beta\Delta t}) + ikx_0 e^{-\beta\Delta t}}$$

$$= \int_{-\infty}^{\infty} e^{-ikx} e^{-2k^2} e^{-\delta k} dk = \int_{-\infty}^{\infty} dk e^{-ikx} \ddot{G}(k)\hat{F}(k) = \mathfrak{F}^{-1}\left[\hat{G}(k)\hat{F}(k)\right]$$

$$= \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} \mathfrak{F}^{-1}\left[\hat{G}(k)\right]\mathfrak{F}^{-1}\left[\hat{F}(k)\right] dk$$

By convolution. FINALLY

$$P(x, t|x_0, t_0) = \frac{1}{\sqrt{2\pi(1 - e^{-2\beta t})\sigma^2}} e^{-\frac{(x - x_0 e\text{^}(-\beta\Delta t)^2}{2\sigma^2(1 - e^{-2\beta\Delta t})}}.$$

(Abstraction for the sake of efficient exploration is fine if it has good foundations). Thanks. A. Alaniz.

Some final thoughts. We humans have compiled a pretty good smallish set of engineering manuals that span from the minutia of particles and fields all the way to the observable cosmos and all things in between that we can engineer and/or model from space stations to play stations and DNA logic gates, stars and nuclear devices, weather patterns and stock markets. For me these notes mark my attempt to understand as much of observational phenomenology as possible with the least amount of mathematical machinery that maximizes the unity of ideas while minimizing prescriptive, voodoo mathematics. Without, however, having a background covering the gamut of nuclear weapons design physics from high explosive physics to radiation transport and the numerical methods techniques inherent in simulating these kinds of weapons, my experience at DARHT, in molecular dynamics and financial physics ranging from natural gas derivatives to climatology, and much other sundry experience, these notes would be no more to me than a meatless skeleton devoid of any intuition that comes from painstaking struggle and eventual hands on familiarity. At a minimum, core graduate physics courses supply sufficient foundations.

The extremes of our knowledge small and large apply far beyond the everyday engineering world. We don't need to know about quarks to make refrigerators or lunar bases or to design genetic corrections. Isn't a smallish codification of observation what many physicists have dreamt of? In this sense there is a great measure of completeness for me in these notes bounding what we know, but I also know that the extremes of our knowledge are vanishingly small. The Lie groups that we use to codify known particles and fields can be embedded in far larger, far richer Lie groups allowing us to theorize about possible extensions of our universe, or other possible universes. The border separating what we know, moreover, from what might be possible is our empirical knowledge, and it's getting ever harder to generate fundamental data. The LHC is a monster in terms of international cost. Progress will be made, but is it asymptotically bound to practical limits from cost to technology? Indeed progress might also be bound by physics itself, if, say, we build a big enough machine to start making little, rapidly evaporating black holes.

Beyond the semi-intuitive theoretical extensions of our Standard Model to larger, higher-dimensional Lie groups, (beyond this set of "core" notes) there are other relatively intuitive extensions to consider. There is the text by M. Nakahara, "Geometry, Topology, and Physics", 2^{nd} ed. It reaches out "linearly" to generalizations of the covariant derivative for example, but it also tickles the edge of chaos where theoretical physics and pure mathematics merge into the wild, wild west. "The Knot Book, An Elementary Introduction to the Mathematical Theory of Knots," C. C. Adams, treating mostly two-dimensional knots in our usual three-dimensional space seems to me to be somewhat like genetics. How many different genera has nature concocted from a set of merely four letters? How far will we humans and our descendants be able to take this four letter alphabet beyond cats who glow with the genes of bioluminescent sea life? How many different classification schemes will lead to light in knot theory? By surfaces, by simplexes, by crossing numbers, by Conway's notation, by polynomials, etcetera, etcetera, etcetera? How much of this is mathematics? How much of this stuff at and beyond three dimensions is actually taken advantage of by physics? Read up on, "The Jones polynomial and quantum field theory," by E. Witten, Communications in mathematical physics, 121, 351-399 (1989). How many ways can string theorists fold higher dimensional manifolds? The degrees of freedom of mathematics seem to greatly exceed the explanatory needs of physics running up as it is against brick walls in costs and potential deserts of no new physics beyond the Standard Model up to some unknown energy threshold.