Optimal shapes for Gears

Warren D. Smith*
Email: wds@research.NJ.NEC.COM

date = December 22, 1993

Abstract — **Gear theory is re-examined and we find optimal shapes for gears. As optimality criteria, we allow: (1) minimal frictional losses (highest efficiency) assuming linear law of friction *or* (2) uniform maximum stress (it will wear out slowly and last the longest) assuming Hertzian contacts *or* (3) uniform maximal temperature, assuming we are in a high power limit in which all heat is removed by the lubricant. Some other criteria which have been used before are that (4) mis-spacing the gears still yields perfect gear action with the desired speed ratio *or* (5) minimal vibration in the low friction limit. Both (4) and (5) lead to "involute gears" which are the standard in engineering practice. Criteria 1, 2, and 3 lead to apparently new gear-tooth forms. We manage to describe these curves with ordinary differential equations (ODEs), and for each of these 3 criteria we find the ODE for both spur (cylinder) and bevel (conical) gears, i.e. a total of 6 ODEs.**

Keywords — minimum wear, uniform stress, minimum temperature, maximal efficiency, minimal frictional loss, optimal shape, form, structure, spur gears, bevel gears.

Contents

*NEC Research Institute, 4 Independence Way Princeton NJ 08540 USA

1 Introduction

First, we will exposit, re-examine, and formalize the basics of gear theory.

Second, we will obtain the following results about "optimal" gear shapes. As optimality criteria, we allow: (1) minimal frictional losses (highest efficiency) assuming linear law of friction *or* (2) uniform maximum stress (it will wear out slowly and last the longest) assuming Hertzian contacts *or* (3) uniform maximal temperature, assuming we are in a high power limit in which all heat is removed by the lubricant. Some other criteria which have been used before are that (4) mis-spacing the gears still yields perfect gear action with the desired speed ratio *or* (5) minimal vibration in the low friction limit. Both (4) and (5) lead to "involute gears," which are the standard in engineering practice, see §6 and §7. Criteria 1, 2, and 3 lead to apparently new gear-tooth forms. We manage to describe these curves with ordinary differential equations (ODEs), and for each of these 3 criteria we find the ODE for both spur (cylinder) and bevel (conical) gears, i.e. a total of 6 ODEs.

In §20 we show how to modify our optimal shapes if there are multiple points of contact.

The min-friction (theorem 10) and uniform temperature (theorem 11) gear-tooth shapes we derive are both "Don Juan,", that is, if one of these gears is mated with another "conjugate" gear, *no matter of what pitch radius,* the result will be an optimal matched pair. Don Juan gears are not as good as having a complete interchangeable family of gears, any member mateable with any other, and with the result always being an optimal matched pair, but this is still a fairly useful and strong statement. (The uniform max-stress gears are apparently not Don Juan.)

To get these new results we will be assuming that the frictional force obeys the usual linear "law"

$$F_{\text{friction}} = \mu F_{\text{normal}}. \qquad (1)$$

where the "coefficient of friction" μ is a positive constant. In some cases we will be assuming that $\mu \ll 1$. In practice μ is somewhat nonconstant and depends on sliding speed. To the extent that such nonlinear effects occur, our analysis will not be valid[1]. We assume that our gears have elastic moduli that are much larger than the stresses they will encounter in use, and that the material they are made of has much larger speed of sound than the speeds of motion of the gears in use.

2 The most general kind of gears with extended contact and "no" sliding.

In this paper, we will only consider gears in which there is "no" sliding contact. Now actually, as will later become clear (cf. §10) all gears that are rigid bodies and which transmit force normally to the contact, *must* feature sliding contact. When we said "no" sliding above, we really meant that in the limit where the teeth have zero height and the gear reduces to a smooth surface, that the action of the two meshed gears reduces to pure rolling contact.

So, to be precise, we are now going to find the most general class of *surfaces* such that (1) pure rotation of each at constant angular rate induces pure rolling contact and (2) the contact is over an extended curve (i.e. is not a point).

First: the surfaces must be bodies of revolution, for if not, the angular speed ratios of the two gears could not be constant.

Second: the radii (distance to axial line) of the two gears, at any particular point of contact, must be in *fixed* ratio.

Given any two axial lines (not necessarily parallel or perpendicular) in 3-space and a desired angular speed ratio, these two conditions suffice to uniquely determine the surfaces of the two gears. Degenerate cases occur where the

[1] However, the "uniform stress" gears of theorem 8 will still have very uniform stress, in spite of any nonconstancy of μ, provided $\mu \ll 1$. This is since almost all of the stress will arise from normal force in this case.

shaft lines are coplanar. In the degenerate case where the shafts are parallel, the answer is two cylinders. In the degenerate case when the shafts are intersecting lines, the answer is two cones with common apex. In the general case (non-coplanar shafts), the answer is a hyperboloid of revolution.

Specifically, let the two shaft lines have closest points A and B. Consider a point C which divides line segment $\overline{AB}$ into parts with the desired ratio. Similarly, let the points $A(s)$ and $B(s)$ be the points of shafts 1 and 2 at distance s from $A = A(0)$ and $B = B(0)$, respectively, and let $C(s)$ be the point which divides line segment $\overline{A(s)B(s)}$ in the desired ratio. The points $C(s)$ form a line called the "generatrix." This line must lie on the gear surfaces. In fact, rotating it about either shaft produces a hyperboloid of revolution. The radius $R(s)$ of the hyperboloid as a function of distance s along the shaft obeys an equation of the form $R(s) = \sqrt{cs^2 + R(0)^2}$ where c is a positive constant. The curve of contact between the two gears is in fact a line, in fact it is the generatrix.

Such hyperboloid gears can be built and line teeth may be cut along the generatrices. They are very beautiful to watch in action. But anyway, we observe that sliding in such hyperboloid gears is unavoidable. Why? Consider the two minimal circles of the gears. These circles touch at a point but their axes are not coplanar. Hence, no matter at what angular rate the circles rotate, there must be sliding action across the plane of one of the circles. This sliding action is along the generatrices which is why hyperboloid gears can function (albeit with sliding along the teeth).

The conclusion is the following theorem (which was part of the folklore, but I had not previously seen a formal proof.

Theorem 1 *The only kinds of surfaces such that (1) pure rotation of each at constant angular rate induces pure rolling contact and (2) the contact is over an extended curve (i.e. is not a point) are: Cones with common apex, and parallel cylinders.*

Note: By imposing the curve-contact condition, we have excluded "crossed helix" gears. Such gears with point contacts wear out quickly. By imposing no-sliding, we have excluded from consideration "worm" drives, crossed helices (again), and "nuts and bolts." (The latter also violate the pure-rotation condition.) All of these excluded gear types are comparatively innefficient since they have comparatively large frictional losses. Typical efficiencies [5] achieved in practice with bevel (cone), and spur (cylinder) gears are 97-99.7%. Hyperboloid, worm gears, and crossed helices tend to have much smaller efficiencies, 50-95%.

All calculations with Bevel gears (cones) work exactly the same way that calculations with spur gears (cylinders) work, except that instead of using Euclidean geometry in the plane of rotation, one works in the nonEuclidean geometry on the surface of a sphere whose center is the intersection of the axes of rotation.

Often spur gears are used with teeth cut, not parallel to the cylinder's axis, but instead at a slight angle (helix). (A similar trick is also used with bevel gears.) This trick is purely intended to average over phase resulting in (1) smoother operation and (2) avoidance of weak points in the phase vs strength graph to get higher load capability. (And one often uses two opposed helices, one left handed, one right handed, on the same gear to avoid any net side thrust.) If the helical teeth are cut correctly, all our analysis is left unaffected, since the helix may be thought of as an infinite number of infinitesimally thin adjacent ordinary spur gears, each rotated a little with respect to its neighbor.

3 Gear terminology

- Bevel gears have teeth cut into conical surfaces and their shafts are intersecting lines. Spur gears have cylindrical surfaces and their shafts are parallel lines.
- Pitch circle: a circle having the "effective radius" of the gear (used in figuring speed ratios). This is also called the pitch radius. Two properly meshed gears have tangent pitch circles. For bevel gears the pitch radius is an angle (and one should imagine the pitch circles as being drawn on the sphere of rotation), for spur gears it is a length (and one should imagine the pitch circles as being drawn on the plane of rotation).
- Sphere of rotation: For bevel gears, the sphere of radius 1 whose center is the point at the intersection of the two shafts. (For spur gears, there is instead a plane of rotation, perpendicular to the two shafts.)
- Addendum: portion of tooth outside pitch circle.
- Dedendum: portion of tooth inside pitch circle.
- Internal vs External: An internal spur gear has teeth cut into the inner circle boundary of an annular disk. An external gear has teeth cut into the boundary of a circular disk.
- Point of contact: two meshed gears in proper operation will have at least one active point of contact at all times. (Actually reversible gears have at least two, but only one of these is active, i.e. has positive pressure.)

Here and below we are working on the sphere/plane of rotation. Of course in three dimensions the "point" of contact is in fact a line segment.

- Normal: There is a common normal direction at the point of contact.
- Centerline: line through the centers of the two gears.
- Pitch point: intersection of pitch circles and centerline.
- Rack: a gear of infinite radius. (For bevel gears, the analog of a rack is a flat disk [cone with apex angle 180°] with teeth cut radially into its top surface.)
- Shape of gear curve: will be specified in this paper by $r(\theta)$ in polar coordinates in the plane of rotation. For bevel gears we use the same method except that r is an angular (geodesic on the sphere of rotation) distance. We will also use $\psi(\theta)$ to denote the angle between a radial line (geodesic) and the curve of the gear. We have that

$$\cot\psi = \mp r'/r\,,\quad \cos^2\psi = \frac{r'^2}{r'^2+r^2}\,,\quad \sin^2\psi = \frac{r^2}{r'^2+r^2} \tag{2}$$

in the plane (for spur gears) and

$$\cot\psi = \mp\frac{r'}{\sin r}\,,\quad \cos^2\psi = \frac{r'^2}{r'^2+\sin^2 r}\,,\quad \sin^2\psi = \frac{\sin^2 r}{r'^2+\sin^2 r} \tag{3}$$

on the sphere (for bevel gears).

4 History of gears

It is not known who invented gears and when. The first known instance: Gears were used for a water clock built in 280 BC, according to a book by Marcus Vitruvius (Roman engineer) in 25 BC. (*The ten books of architecture,* available in English translation as a Dover reprint.) It is possible that worm gears were used by Archimedes ($\approx$ 287-212BC) in building military catapults. However, Archimedes is known to have invented the compound pulley, which also could have served this purpose, perhaps more simply. So perhaps gears were invented $\approx$ 300BC.

According to [9] in the year 65 ± 15 BC, a large ship was wrecked which carried a sun-moon-planets computing machine containing > 20 bronze gears including crown, spur, and planetary.

Girard Desargues (1591-1661) invented and made cycloidal gears, the first system of interchangeable spur gears.

Phillipe de la Hire (1640-1718), a student or associate of Desargues's, invented and made involute gears and wrote *Traité des épicycloïdes et de leur usage dans les méchaniques* in 1694.

Others who studied gears mathematically include Charles Camus (1699-1768), Leonhard Euler (1707-1783), Jacques Savary ($\approx$1760), and Edward Sang ($\approx$1840). Known theory was exposited in the books Th. Olivier: *Theorie Geometriques des engrenages*, Paris 1842, and Franz Reuleaux: *Theoretische kinematik,* Vieweg, Braunschweig 1875 (there is an English translation of this book by A.B.W.Kennedy: *The kinematics of machinery,* London 1876, Dover reprint 1963). I do not know what these people did, and I have been unable to obtain these books.

The contemporary book most renowned for its theoretical content is Buckingham's [2]. Others include [8], [5].

Many gears were standardized by the ASME, AGMA, and/or ANSI in 1940-1970, and since that time, fundamental progress in gear design has virtually ceased[2].

5 Fundamental law of gear action

To preserve constant ratio of angular velocities[3], gears have to have the property that a line drawn in the normal direction through the point of contact, called *the normal line*, must cross the centerline at a fixed location, namely the same place where the pitch circles are tangent, namely the *pitch point.*

[2] An ingenious invention by C.W. Musser in the 1950's (US pats. 3117763, 3088333, 3435705) which is the exception to this rule is the "harmonic drive," a kind of gear suitable for very high reduction ratios and whose operation depends on flexible materials. It can achieve reduction ratios from 50-320:1, with efficiencies of 80-90%, at high torque in a light and compact assembly. More information may be obtained from Harmonic Drive Technologies, Teijin Seiki Boston Inc., 247 Lynnfield St., Peabody MA 01960 and also the Encyclopedia Brittanica article on "harmonic drive."

[3] It is also possible to require constant ratios of torque. This is equivalent to having constant ratio of angular velocities if and only if the coefficient μ of friction is *zero.* The law of constant torque ratio action would be as follows. "A line, the *modified line of action*, through the point of contact and at angle $\arctan(\mu)$ to the normal line, must contain the pitch point." The involute of a circle is a curve that plays a very important role in the theory of spur gears. The curves playing the same role for gears with constant torque ratio obey the equation

$$\theta_0 \pm \theta = \frac{1}{\mu}\log(1-\mu\sqrt{r^2/c^2-1}) + \mathrm{arcsec}(r/c). \tag{4}$$

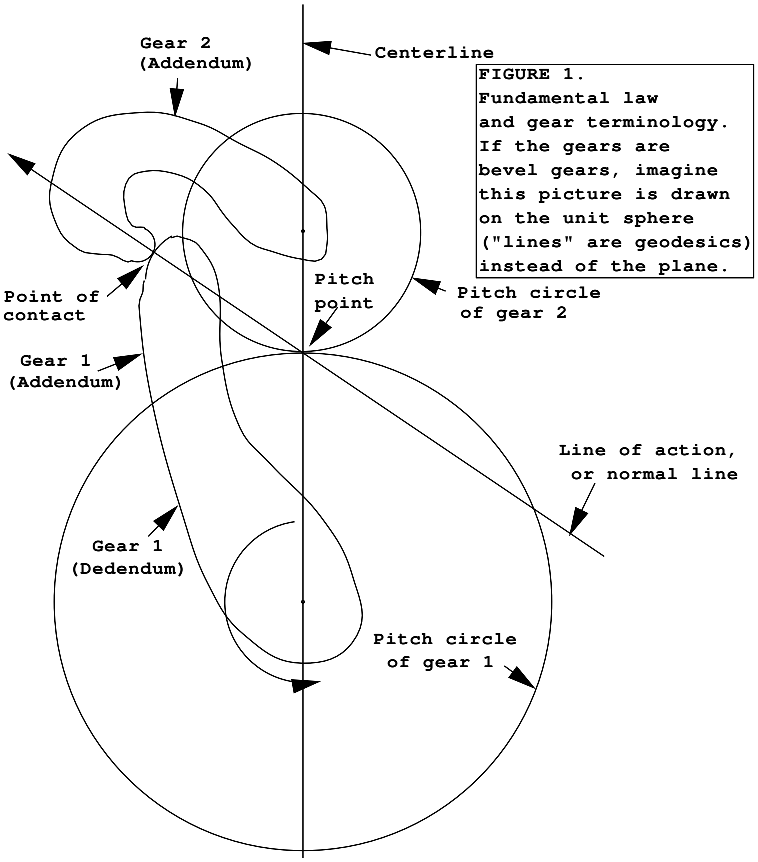

Figure 1:

For bevel gears, the fundamental law states that the normal line (which is now actually no longer a line, but in fact a geodesic, drawn on the sphere of rotation, through the point of contact and normal to it) must contain the pitch point. Viewed three dimensionally, this statement is that the normal *plane* (defined by the normal "line" and the point of intersection of the axes) must contain the pitch *line* (line of tangency of the pitch cones).

6 The two kinds of gears which have been commonly used

An infinite number of gear tooth shapes satisfy the fundamental law. One wishes to find the one which optimizes something.

Two kinds of gears have been extensively used.

Cycloid gears: An imaginary "generating circle" of standard size is supposed to be rolling along the inside or outside of the gear's pitch circle. Indeed, as two meshed gears rotate, the two generating circles also rotate, with all 4 circles having centers fixed on the centerline. A point on the generating circle traces out the tooth on the rotating gear. Then you switch to the other generating circle when the tracing point hits the pitch point. The rack has teeth whose addenda and dedenda are equal cycloids. During operation, the point of contact in cycloid gears travels on a continuous curve, namely it is the "figure 8" that is the union of the two generating circles. The angle of the normal in cycloid gears changes continuously, indeed it rotates at a constant angular rate which is half the rate of rotation of the imaginary generating circles.

Involute gears: These are the standard in present day engineering. Cycloid gears used to be usual but they are now obsolete and I think are only still used in clockwork [which has high precision, and needs low friction] because they supposedly have less friction than involute gears. The *involute of a circle* is the curve traced out by a point of a taut string unwinding from the circle. (Equation: in the plane [spur gears] $\theta_0 \pm \theta = \sqrt{(r/r_0)^2 - 1} - \arccos(r_0/r)$.

The assumption of constant angular velocity ratio is a much more realistic and desirable one, in most practical situations, than the assumption of constant torque ratio. In this paper we will often make the assumption that $\mu \ll 1$ in which case the two laws are the same and force (which is transmitted along the modified line of action) will always be transmitted along the normal line. Entirely generally, any mechanical linkage which, when frictionless, would transmit constant force ratio, will, even with friction, transmit constant velocity ratio.

On the sphere [bevel gears],

$$\theta_0 \pm \theta = \frac{1}{\sin r_0} \arccos(\frac{\cos r}{\cos r_0}) - \arccos(\frac{\tan r_0}{\tan r}) . \quad) \tag{5}$$

One imagines a string simultaneously winding onto a circle concentric with, but with a radius a standard constant factor smaller than, one gear's pitch circle and unwinding from a similarly defined circle on the other gear. An imaginary *dual* string goes the other way (forming an X with the first string). A point on the string traces out the teeth and follows the point of contact. You switch to the other string, to do the other side of the tooth. The curve followed by the point of contact is a line segment. It is traversed at constant speed. It is not a continuous closed curve. The active surfaces (sides) of the teeth are involutes. The inactive surfaces are simply the base and tip circles and if properly made, should never touch anything. The rack in the involute system has teeth which are trapezoids (the involute degenerates to a straight line).

Also, in both systems, internal gears are the set-complement of external gears.

The fact that involute gears involve a trapezoid rack is an advantage of the involute system since it makes racks easy to manufacture, and also makes gears easy to make since you can cut them using a hob. A hob is a trapezoid rack, cartesian product in the 3rd dimension with cutters. By rotation of the gear and translation of the hob by the appropriate amounts to simulate a small time interval of gear-rack action, then a cutting stroke, then another simulated timestep, then another cutting stroke, etc, involute gears will be cut. I should point out that in practice one would want to round the corners of the trapezoids since sharp re-entrant corners tend to cause huge stress amplification and cracks to form.

Notice both these gear types are general purpose: they do not need to be made as matched pairs. They will work with any other size gear (subject to integer number of teeth...) as the other member of the pair.

7 Known optimal properties of involute gears

Known unique advantages of involute gears (and the reasons why they could be said to be "optimal") are as follows.

The normal always has a fixed orientation. So the normal line is a fixed line. In fact, it is the line of that imaginary string. So in addition to the transmitted torque being constant (if $\mu \ll 1$), the transmitted *force* is constant. Involute gears are the only thing with this property. Presumably, this property leads to *minimal vibration.*

Also, while cycloid gears have to be very precisely spaced, involute gears can have a little error in the spacing and it will still work well; in fact, *mis-spaced involute gears are still matched involute gears* (just with a slightly different force-transmission angle and different pitch circles) and *still having the same pitch ratio.* All other types of gears have to have preselected pitch circles. This is a tremendous practical advantage.

We have already mentioned how easy involute gears and racks are to manufacture using trapezoid hobs.

The discontinuous path followed by point of contact in involute gears may be a disadvantage from point of view of sealing out dirt, but it can also be an advantage since it allows more than 1 point of contact to exist (time averaged contact number has been > 1.4 in some involute gears which have been made) perhaps increasing the allowable load.

Cycloid gears still work even if they are highly exaggerated (i.e. "gears" with only one or two "teeth"). But I do not know of any optimality properties of cycloid gears. The pre-eminence they enjoyed at one time seems to have been solely due to the fact that they were the first family of interchangeable gears to be discovered which obeyed the fundamental law.

8 The conjugate of a gear

In this section we describe how to generate the unique gear, with some desired pitch radius, that is "conjugate to" (i.e. mates with) any given gear.

We will solve this problem both in the plane and on the surface of a sphere. The former problem is relevant for spur gears, the latter for bevel gears.

Call the given gear "gear 2," and its conjugate, that we are trying to find, "gear 1." We assume that the form $r_2(\theta_2)$ (and hence also $\psi_2(\theta_2)$) of gear 2 is known. Consult figure 2. Gear j has pitch radius R_j and is rotating at constant angular velocity ω_j. r_j, θ_j, and ψ_j describe the curve of gear j, and all three of these quantities depend upon the time t at which that point becomes the point of contact, although (we caution the reader) we will not always state such dependence explicitly in our formulas. P is the point of contact. We will first consider the more complicated *bevel* gear problem where all gear curves are drawn on a *sphere* of unit radius. For bevel gears, R_j are angles and $\omega_2 \sin R_2 = \omega_1 \sin R_1 = V$. For spur gears, R_j are lengths and $\omega_2 R_2 = \omega_1 R_1 = V$, where V is the pitch circle velocity.

From [4], page 33, we see that

$$\cot R_2 \sin r_2 = \cos r_2 \cos(\omega_2 t - \theta_2) + \sin(\omega_2 t - \theta_2) \tan \psi_2 \tag{6}$$

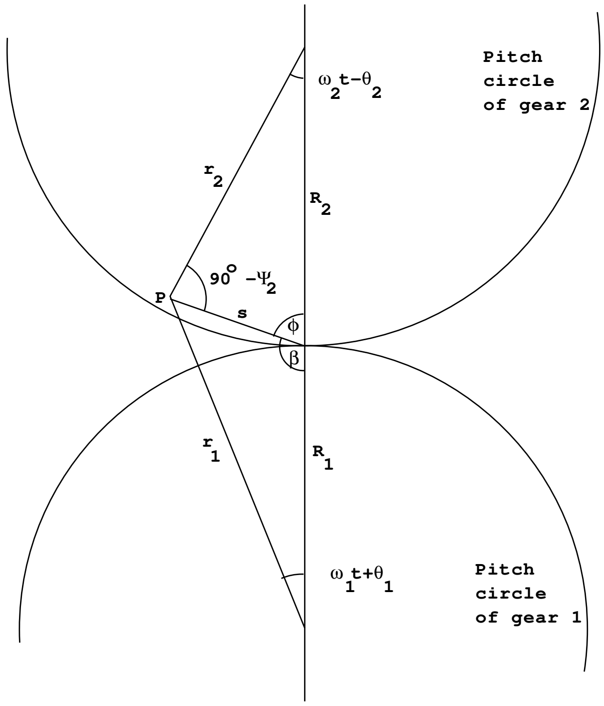

Figure 2:

which allows us to solve for $\theta_2(t)$, although in general (since ψ_2 and r_2 depend on θ_2 in possibly ugly [although known] ways) this solution is not going to have an explicit closed form and will only be available implicitly. (In the exceptional planar example of the "pin tooth gear," see below, there is a closed form.) And then of course since the curve of gear 2, namely $r_2(\theta_2)$, is known we know the location of the point P of contact as a function of time t. It is convenient to find this out explicitly by solving for $s(t)$ and $\phi(t)$. By the spherical law of sines we find

$$\sin s = \frac{\sin(\omega_2 t - \theta_2)}{\cos\psi_2} \sin R_2 \, , \quad \sin\phi = \frac{\sin r_2}{\sin R_2} \cos\psi_2 . \tag{7}$$

(Many alternative expressions are also possible.) Finally, we wish to find the curve $r_1(\theta_1)$ of the conjugate gear. Actually we will find a parametric version of the same thing, namely $r_1(t)$ and $\theta_1(t)$, which has the added bonus that it describes the location of the point of contact. This is accomplished by ([4] pages 28,33)

$$\cos r_1 = \cos R_1 \cos s + \sin R_1 \sin s \cos\phi \, , \quad \tan(\omega_1 t + \theta_1) = \frac{\sin\phi}{\cot s \sin R_1 \; + \; \cos R_1 \cos\phi} \tag{8}$$

Now the analagous but simpler development for a *spur* gear (all gear curves drawn in a *plane*) is

$$\frac{r_2}{R_2} = \cos(\omega_2 t - \theta_2) + \sin(\omega_2 t - \theta_2) \tan\psi_2 \tag{9}$$

allowing one to solve for $\theta_2(t)$; then the polar coordinates $s(t)$ and $\phi(t)$ of the point P of contact are given by

$$s = \frac{\sin(\omega_2 t - \theta_2)}{\cos\psi_2} R_2 \, , \quad \sin\phi = \frac{r_2}{R_2} \cos\psi_2 . \tag{10}$$

Then the curve of the conjugate gear is described parametrically by

$$(r_1)^2 = s^2 + R_1^2 - 2R_1 s \cos\phi \, , \quad \tan(\omega_1 t + \theta_1) = \frac{s \sin\phi}{R_1 \; + \; s\cos\phi} . \tag{11}$$

As an example of this sort of procedure in action, we will now derive the curve of the gear of pitch radius R_1 that is conjugate to a "pin-tooth" gear, of pitch radius R_2, whose profile is a circle (or several circles), or "pin," of radius B, centered at a distance A from the center of the gear. (This calculation was carried out in [2], but not analytically.) Consult figure 3.

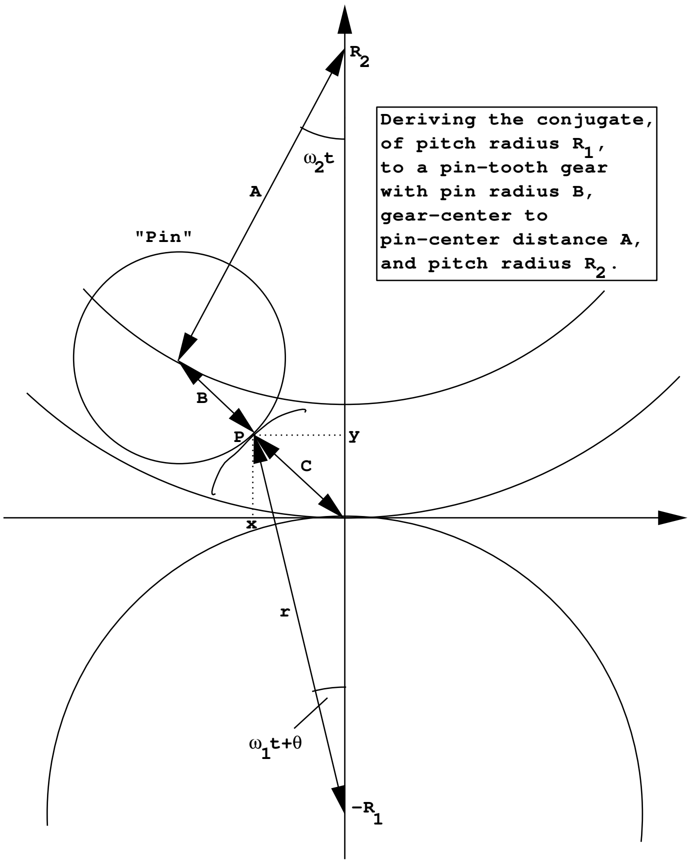

Figure 3:

For simplicity set $\omega_1 = 1$ and $\omega_2 = R_1/R_2$. We can see by the law of cosines that

$$C(t) = \sqrt{A^2 + R_2^2 - 2AR_2 \cos\frac{R_1 t}{R_2}} - B. \tag{12}$$

Hence the cartesian coordinates x, y of the point P of contact are

$$(x(t), y(t)) = \frac{C(t)}{C(t)+B}(-A\sin\frac{R_1 t}{R_2}\ ,\ R_2 - A\cos\frac{R_1 t}{R_2}). \tag{13}$$

So, since the curve of gear 2 in terms of the cartesian coordinates of P is given in general by

$$r = \sqrt{x(t)^2 + [y(t)+R_1]^2}\,, \quad \theta = \arg(y(t) + R_1 - ix(t)) - \omega_1 t, \tag{14}$$

in our case the curve $r(\theta)$ of gear 1 is described parametrically by

$$r = \sqrt{(\frac{C(t)A}{C(t)+B})^2 + (R_1 + \frac{C(t)}{C(t)+B}R_2)^2 - 2(R_1 + R_2\frac{C(t)}{C(t)+B})\frac{C(t)A}{C(t)+B}\cos\frac{R_1 t}{R_2}}, \tag{15}$$

$$\theta = \arg\left(\frac{C(t)+B}{C(t)}R_1 + R_2 - A\cos\frac{R_1 t}{R_2} + iA\sin\frac{R_1 t}{R_2}\right) - t \tag{16}$$

where $C(t)$ is given in (EQ 12).

9 Interchangeability, matched pairs, and Don Juan gears

For reasons of cost, one generally wishes for gears to be *self-conjugate* so they will mesh with copies of themselves and run in either direction, and more generally that all gears be interchangeable – any gear will mesh with any other. We have already observed that the cycloid and involute gears form an interchangeable system. This might at first be thought to be miraculous, but as we will see, interchangeability is hardly a restriction at all.

We first observe that self-conjugacy for a *rack* is trivial: it is merely the condition that the set-complement and the mirror image of a rack should be congruent to the rack itself. That is, the profile of the rack should be a periodic function with respect to the pitch line having both odd symmetry (about the point where the tooth profile crosses the pitch line) and also mirror symmetry about the middle of the tooth.

Then, gears of any radius (ignoring issues of tooth-count integrality) may be defined as the conjugates of this rack. Two important folklore theorems (which have been referred to as "Sang's theorems," and which also hold for bevel gears, if interpreted appropriately) are:

Theorem 2 *Any two gears defined by an odd and even symmetric rack, will run in either direction when meshed with each other, and their action will obey the fundamental law of §5. (Caution. This theorem is not intended to rule out the possibility that the teeth may overlap elsewhere from "the" point of contact; it is purely intended to be a "local" statement.)*

Theorem 3 *If* A,B,C,D *are gears (and/or racks), and* $A \hookrightarrow B$, $B \hookrightarrow C$, $C \hookrightarrow D$, *then* $A \hookrightarrow D$, *where* "$X \hookrightarrow Y$" *is to be read* "X *is conjugate to* Y."

Proof. Imagine that all four gears A,B,C,D are placed on the same centerline with the same pitch point. (Gears C and A and gears B and D, will overlap mentally.) Then it follows from the fundamental law (§5) that the point of contact between A and B is the same as between B and C, is the same as between C and D. Theorem 3 is now proven and theorem 2 follows. □

On the other hand, if we want very much for a particular gear to optimize some property, then we do not care about interchangeability and self-conjugacy[4].

Most of the rest of this paper will be concerned with that precise situation; we will focus on one particular gear and demand that it optimize some property when matched with a conjugate gear. On the other hand, we make no claim that the resulting gear will generally be self-conjugate.

In the present age of computer controlled cutting tools, optimized matched pairs may not be unrealistic.

Besides, theorems 10 and 11 will give us something more than just optimized matched pairs; they in fact generate "Don Juan" gears such that, when they are mated with *any* conjugate, no matter of what radius, the result is an optimal matched pair. Also, by combining (EQ 20) and (EQ 47), we could write the differential equation for a gear which would make the rate $\mu F_{\text{normal}} V_{\text{sliding}}$ of frictional heat production constant. I do not know why one would want a gear with this property, but, the resulting shape would be independent of the radius R_2 of the other gear (except for a constant factor), hence this kind of gear would also be Don Juan.

On the other hand, theorem 8 will only yield optimal matched pairs and apparently will not yield Don Juan gears.

10 Sliding friction cannot be avoided

The sliding velocity at the point of contact may be written

$$V_{\text{sliding}} = (\frac{r_1 \sin\psi_1}{R_1} - \frac{r_2 \sin\psi_2}{R_2})V \tag{17}$$

in the plane (for spur gears) and

$$V_{\text{sliding}} = (\frac{\sin r_1 \sin\psi_1}{\sin R_1} - \frac{\sin r_2 \sin\psi_2}{\sin R_2})V \tag{18}$$

on the sphere (for bevel gears), where V is the velocity of the pitch circles. There is a nice way to rewrite this formula in such a way that the only quantities in it are lengths, and none of them are angles. We first observe that $\sin r_i \sin\psi_i / \sin R_i = \tan\psi_i \sin\phi$ by the spherical law of sines, where ϕ is the angle between the centerline and the normal line. Let s be the distance from the pitch point to the point of contact. We then solve for angles ψ_i in terms of the sides R_i, ϕ, and s using the "4 adjacent parts"la from spherical trigonometry. The final result is

$$V_{\text{sliding}} = (\frac{1}{\tan R_1} + \frac{1}{\tan R_2})V \sin s \tag{19}$$

for bevel gears, which reduces to

$$V_{\text{sliding}} = (\frac{1}{R_1} + \frac{1}{R_2})Vs \tag{20}$$

[4] Also, if we knew ahead of time that a gear-pair would always be running in a particular direction with one driving the other, we could drop the reversibility (even symmetry) requirement.

for spur gears.

Incidentally, if in the above it is desired to express s in terms of the curve of gear 1 alone, then ([4] page 28)

$$\cos R_1 = \cos r_1 \cos s + \sin r_1 \sin s \sin \psi_1 \tag{21}$$

for bevel gears, so that $\sin s$ is the solution of a quadratic equation:

$$\sin s = \frac{\cos R_1 \sin \psi_1 \sin r_1 \pm \cos r_1 \sqrt{\sin^2 R_1 - \cos^2 \psi_1 \sin^2 r_1}}{\sin^2 \psi_1 \sin^2 r_1 + \cos^2 r_1}. \tag{22}$$

For spur gears this is

$$s = r_1 \sin \psi_1 \pm \sqrt{R_1^2 - (r_1 \cos \psi_1)^2}\,. \tag{23}$$

In general I'm going to stick to the upper sign both here and in (EQ 2) (EQ 3).

In principle all sliding friction could be avoided by using a pin-tooth gear with the pins having ball bearings at their ends. However, sliding friction in rigid-body gears is unavoidable:

Theorem 4 *In any spur or bevel gears which are rigid bodies obeying the fundamental law (§5), and where plain circular wheels are not allowed, sliding friction is unavoidable. The velocity of sliding is given by (EQ 20) for spur gears and (EQ 19) for bevel gears.*

Proof. We've proven (EQ 20) and (EQ 19) above. $V_{\text{sliding}} = 0$ is possible only if the point of contact is always identical with the pitch point, which is possible only for plain circular wheels. □

Note that the direction of sliding changes (and $V_{\text{sliding}} = 0$) when the point of contact hits the pitch point.

Rolling contact (smooth circles) and/or belts and pulleys could be used if one really wanted to eliminate sliding friction. But you wouldn't have much allowable torque then; the whole point of gears as opposed to friction wheels is that they can transmit much more torque.

11 Is friction linear?

And what about with lubrication?

The answer is: yes, experimentally, friction is approximately linear, as in (EQ 1), and μ is approximately constant, and even with lubrication. I recommend the introductory discussion in [1].

According to experiments by S.J.Dokos [3] cited in [1], μ for dry steel on dry steel gradually falls from 0.53 at low speed to 0.18 at high speed as speed is increased by a factor of 10^7. More generally it is agreed that μ *decreases* at high sliding speed for unlubricated surfaces.

For the right kinds of solid surfaces, μ can be quite low even with no lubrication. For example, for teflon-teflon, $\mu = 0.04$. For metals whose surfaces have been treated with stearic acid to form solid soaps, $\mu = 0.05$. On the other hand, degassed metals in vacuum can weld on contact.

For liquid-lubricated plane surfaces with constant liquid thickness, μ is exactly constant, in theory, so long as sliding speeds are slow enough that the preferred solution of the Navier-Stokes equations is steady. In practice, the thickness of the film is nonconstant and the fluid motion can become turbulent, and not necessarily in the lubricating film itself, but elsewhere. Experiments reported by Buckingham [2], pages 404-414, on lubricated cast iron spur gears indicate that an "effective coefficient of friction" μ is dependent on, and *increases* with, pitch circle speed. For example in one experiment, Buckingham (page 404) found $\mu = 0.0112$ at 150ft/min, $\mu = 0.0385$ at 2100 ft/min, and $\mu = 0.0517$ (stationary). Similar experiments reported by Dudley [5] found similar results with μ varying within the range $0.017 \le \mu \le 0.061$ with various different grades of oil and pitch circle speeds from stationary to 5000 ft/min.

According to Buckingham, μ initially drops rapidly as pitch circle speed V increases from stationary to some small speed [10-200 cm/sec] (probably mainly because at very low speeds, contact can occur, but above this critical speed there is always lubricant between) but then gradually increases, eventually apparently proportionally to V^p where $0.51 \le p \le 0.60$. An empirical equation given by Buckingham to "explain" this is $\mu = c_2 \exp(-c_1\sqrt{V}) + c_3\sqrt{V}$. Dudley's experiments seem to indicate less rapid and smaller variations in μ than Buckingham's experiments. Meanwhile, experiments on lubricated *worm* gears reported in [5], page 14-10, indicate that the effective coefficient of friction *drops* with increasing pitch circle velocities (up to 1000 ft/min), apparently asymptotically to V^{-p} where $0.5 < p < 1$.

All in all, friction is little understood. Still, we point out that in Buckingham's, Dudley's, and Dokos's experiments, μ did not greatly vary (only by a factor of 3) over a very large range of nonzero velocities.

At very high power, gears can fail due to heat production being too large for the lubricant to carry it away, hence causing burning or vaporization of the lubricant, or because pressures are high enough to expel the lubricant and force contact. The latter problem can be overcome by the use of "pressure capable" lubricants. (Buckingham p502-505.) The former problem is rare, but could be overcome by using lubricants with high vaporizing temperatures and low reactivity, although in the limit where the lubricant's vaporizing temperature were much higher than the melting point of the gears, this would be of little avail.

12 Curvature and arclength: some useful formulas

Formulae for element ds of arc length:

$$ds = \sqrt{1 + y'(x)^2}\, dx \tag{24}$$

for a curve $y(x)$ expressed in cartesian coordinates, and

$$ds = \sqrt{r'(\theta)^2 + r(\theta)^2}\, d\theta \tag{25}$$

for a curve $r(\theta)$ in polar coordinates, in the plane. On the unit sphere, using "polar" coordinates where r is angular, we have

$$ds = \sqrt{(\sin r)^2 (d\theta)^2 + (dr)^2}. \tag{26}$$

Formulae for radius of curvature ρ:

$$\rho = \frac{(1 + y'^2)^{3/2}}{y''} = \frac{(r^2 + r'^2)^{3/2}}{r^2 - rr'' + 2r'^2} \tag{27}$$

The latter may be more simply written

$$\rho = \frac{r}{[1 - \psi'] \sin \psi} = \frac{\sqrt{r'^2 + r^2}}{1 - \psi'} \tag{28}$$

if $\psi(\theta)$ =angle between the tangent to the curve $r(\theta)$, and the radius vector. The planar formulae above may be derived from the relation $\kappa\rho = 1$ between the curvature κ and the radius of curvature ρ, and the relation $\kappa = d(\theta - \psi)/ds$. On the unit sphere the formula is

$$\tan \rho = \frac{\sin r}{[\cos r - \psi'] \sin \psi} = \frac{\sqrt{\sin^2 r + r'^2}}{\cos r - \psi'} = \frac{(\sin^2 r + r'^2)^{3/2}}{\cos r \sin^2 r - r'' \sin r + 2r'^2 \cos r} \tag{29}$$

This was derived as follows. First, one shows that the relationship between the curvature (turning rate, when moving along the curve with unit velocity) κ is related to the radius of curvature (radius of the osculating circle) ρ by $\kappa = \cot \rho$. To see this, find the curvature of a circle. Second, to write a formula for curvature in spherical polar coordinates (r, θ) , one considers a spherical triangle whose base vertices are two points on the curve at separation ds and whose apex is the center of curvature, and whose apex angle is $d\theta$. Letting the two base angles of the triangle be called A and B, one finds $\kappa = d\psi/ds + (B - A)/ds$ and the second term may be computed from $A + B - \pi + d\theta =$ area where area $= (1 - \cos r)d\theta$. We have chosen the sign of ρ so that $\rho > 0$ for small circles.

13 Radii of curvature of a gear and its mate

We wish to find a relation between the radius c_1 of curvature of a gear and that (c_2) of its conjugate gear at the point of contact $\vec{P} = (\mathrm{Px}, \mathrm{Py})$. **Caution:** note our use of 2-letter variable names such as "Py" in this section.

The centers of the curvature circles are $\vec{Q}_1 = (\mathrm{Qx}_1, \mathrm{Qy}_1)$ and $\vec{Q}_2$. Let the radii of the pitch circles be R_1 and R_2. Let the pitch point be the origin, and the center of gear 1 lies at $(0, -R_1)$, and the center of gear 2 is at $(0, R_2)$. Rotation of gear 1 is counterclockwise. The fundamental law states that the centers $\vec{Q}_i$ of the curvature circles, the point of contact $\vec{P}$, and the pitch point $\vec{0}$ are all collinear, i.e. $\mathrm{Qy}_1/\mathrm{Qx}_1 = \mathrm{Qy}_2/\mathrm{Qx}_2 = \mathrm{Py}/\mathrm{Px}$. We will assume that $\vec{0}$, $\vec{P}$, and $\vec{Q}_2$ are collinear *in this order*, as in figure 1.

Consider the situation a time dt later. Then the rotated versions of $\vec{Q}_1$ and $\vec{Q}_2$ are still collinear. (They are also still separated by distance $c_1 + c_2$, but that will happen automatically.) From $\frac{d}{dt}[\mathrm{Qx}_1 \mathrm{Qy}_2 - \mathrm{Qy}_1 \mathrm{Qx}_2] = 0$ we find

$$(R_1 + R_2)(\mathrm{Qy}_1 \mathrm{Qy}_2 + \mathrm{Qx}_2 \mathrm{Qx}_1) = (\mathrm{Qy}_1 - \mathrm{Qy}_2) R_1 R_2. \tag{30}$$

If we let $\mathrm{Qx}_2/\mathrm{Qx}_1 = \mathrm{Qy}_2/\mathrm{Qy}_1 = z$, then this implies

$$z = \frac{\mathrm{Qy}_1}{(1/R_2 + 1/R_1)(\mathrm{Qy}_1^2 + \mathrm{Qx}_1^2) + \mathrm{Qy}_1} \tag{31}$$

Note that $z = (|\vec{P}| + c_2)/(|\vec{P}| - c_1)$, thus $c_2 = (|\vec{P}| - c_1)z - |\vec{P}|$. Also: $\vec{Q}_1 = \vec{P}(1 - c_1/|\vec{P}|)$. So we have now succeeded in expressing c_2 in terms of $\vec{P}$,c_1,R_1,R_2. Writing $s = |\vec{P}|$ and simplifying with the aid of the computer algebra system MAPLE allows us to express the answer is as follows.

Theorem 5 *Under the fundamental law of gear action, the radius c_1 of curvature of a gear (of pitch radius R_1) at its point of contact with its conjugate (whose pitch radius is R_2) is related to the radius c_2 of curvature of the conjugate gear at the point of contact by*

$$\frac{1}{c_1 \mp s} + \frac{1}{c_2 \pm s} = \frac{1}{\cos\phi}\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{32}$$

If the upper sign is used, gear 1 is on the pitch point's side of the contact. Here ϕ,

$$\cos\phi = \frac{\mathrm{Py}}{s} = \frac{r_1}{R_1}\cos\psi_1, \tag{33}$$

is the angle between the centerline and the normal line, and $s = |\vec{P}|$, $s \mp c_1$, and $s \pm c_2$ are respectively the distances from the pitch point to the point of contact, the center of curvature of gear 1, and the center of curvature of gear 2.

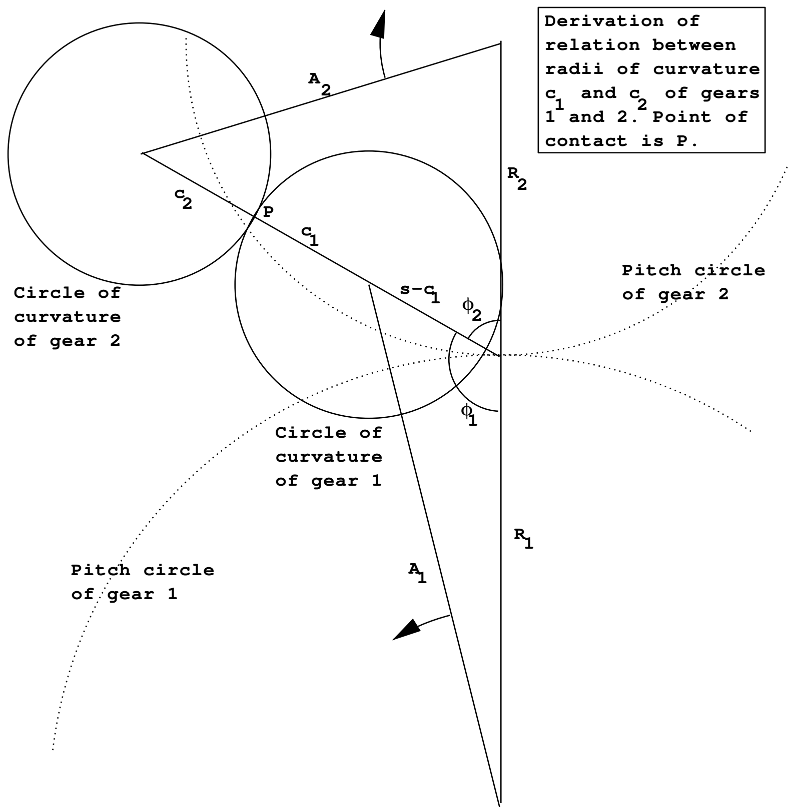

Figure 4:

With the aid of hindsight in the planar case, we will now derive the corresponding relation for bevel gears in a more straightforward manner[5]. (By "radius of curvature" of a curve drawn on the sphere we mean the angular radius of the osculating circle.) Consult figure 4, which is supposed to be drawn on a sphere. We see that as the gears rotate infinitesimally, A_i, R_i, and c_i remain fixed, but s changes. By spherical trigonometry we have

$$\cos A_2 = \cos(s + c_2)\cos R_2 + \sin(s + c_2)\sin R_2 \cos\phi_2 \tag{34}$$

$$\cos A_1 = \cos(s - c_1)\cos R_1 - \sin(s - c_1)\sin R_1 \cos\phi_1 \tag{35}$$

where $\phi_1 + \phi_2 = \pi$ (since the centers of curvature, the point of contact, and the pitch point lie on the same geodesic by the fundamental law) so that $\cos\phi_1 + \cos\phi_2 = 0$. Writing $\frac{d}{ds}(\cos\phi_1 + \cos\phi_2) = 0$ and then using (EQ 34) and (EQ 35) yields

$$\cot R_1 + \cot R_2 = \cot(s + c_2)\cos(\phi_2) + \cot(s - c_1)\cos(\phi_1) \tag{36}$$

leading to the conclusion that

[5] It often pays to derive the plane and spherical formulas separately, even though the former may always be derived from the latter by a limiting process, because by so doing one gets a useful check on one's calculations.

Theorem 6 *Under the fundamental law of gear action, the radius c_1 of curvature of a bevel gear (of angular pitch radius R_1) at its point of contact with its conjugate (whose pitch radius is R_2) is related to the radius c_2 of curvature of the conjugate gear at the point of contact by*

$$\cot(c_1 \mp s) + \cot(c_2 \pm s) = \frac{1}{\cos\phi}(\cot R_1 + \cot R_2) \tag{37}$$

If the upper sign is used, gear 1 is on the pitch point's side of the contact. Here s, $s \mp c_1$, and $s \pm c_2$ are respectively the angular distances from the pitch point to the point of contact, the center of curvature of gear 1, and the center of curvature of gear 2. The angle ϕ between the centerline and the normal line is given by the spherical law of sines

$$\cos\phi = \frac{\sin r_1}{\sin R_1}\cos\psi_1. \tag{38}$$

It turns out that theorems 5 and 6 are not original. (EQ 32) is known as the Euler-Savary equation. The quantity s in them may be obtained in terms of the gear curve using (EQ 23) for spur gears and (EQ 22) for bevel gears.

14 Velocity of the point of contact

In this section we will derive the speed $V_{\text{point of contact}}$ at which the point of contact between conjugate gears 1 and 2 moves, relative to the atoms of gear 2 which are located near the point of contact.

Please do not confuse this with V_{sliding}, which is the speed at which the atoms of gear 1, near the point of contact, are moving relative to the atoms of gear 2. To make the distinction clear, we observe that if gears 1 and 2 were (at some moment) in pure rolling contact, then V_{sliding} would be 0, but $V_{\text{point of contact}}$ would be nonzero, in fact it would be whatever the rolling speed was. It is also possible to construct an example of a pure sliding contact in which $V_{\text{point of contact}} = 0$, while V_{sliding} is nonzero, in fact it would be whatever the sliding speed was. (Consider a rotating automobile wheel failing to move a car off some ice.)

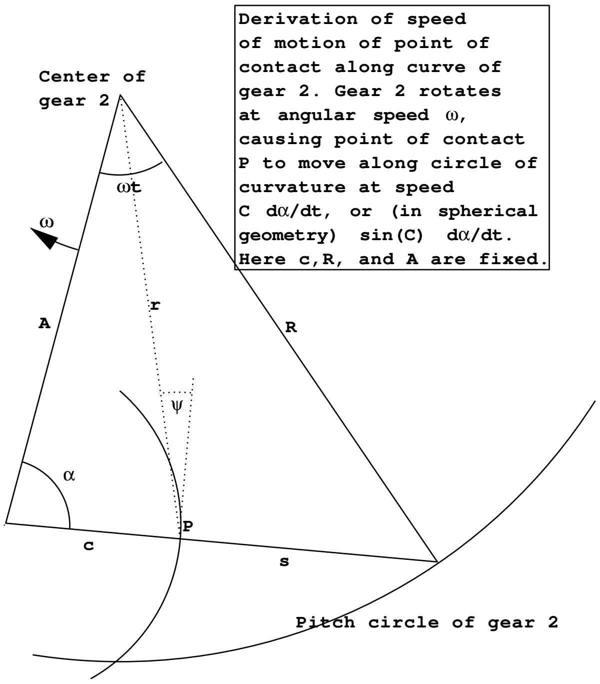

Figure 5:

Consult figure 5.

Let us first regard this figure as being drawn on the unit sphere (bevel gear case). The angle α may be written in terms of R, A, and ωt by standard spherical trigonometry formulas

$$\cot\alpha\sin\omega t + \cos\omega t\cos A = \cot R\sin A \tag{39}$$

which allows us to write

$$V_{\text{point of contact}} = \mp\sin c\frac{d\alpha}{dt}. \tag{40}$$

In this expression, all explicit reference to t may be gotten rid of by the use of the spherical trigonometry formula

$$\cos(c+s) = \cos R \cos A + \sin R \sin A \cos \omega t, \tag{41}$$

and all explicit reference to ω may be gotten rid of by use of $V = \omega \sin R$ where V is the pitch circle velocity. The final result is

$$V_{\text{point of contact}} = \frac{[\cos A - \cos R \cos(c+s)] \sin c}{\sin R \sin^2(c+s)} V. \tag{42}$$

Here, if it is desired, the occurence of $\cos A$ may be gotten rid of too by use of

$$\cos A = \cos r \cos c + \sin r \sin c \sin \psi. \tag{43}$$

Next, if figure 5 is regarded as being drawn in a plane (spur gear case), then (EQ 42) reduces to

$$V_{\text{point of contact}} = \frac{[R^2 + (c+s)^2 - A^2]c}{2R(c+s)^2} V, \tag{44}$$

which by application of the law of cosines $A^2 = r^2 + c^2 - 2rc \sin \psi$ may be converted to

$$V_{\text{point of contact}} = \frac{[R^2 - r^2 + s^2 + 2cs + 2rc \sin \psi]c}{2R(c+s)^2} V. \tag{45}$$

Theorem 7 *The velocity of the point of contact between a gear of pitch radius R and its conjugate gear, is given in terms of the pitch circle velocity V, the radius c of curvature of the gear at the point of contact, the distance s between the point of contact and the pitch point, and the curve $r(\theta)$ of the gear, by (EQ 42) and (EQ 43) for bevel gears and (EQ 45) for spur gears.*

15 Gear that minimizes frictional losses

We want to know what tooth shape minimizes frictional energy loss, assuming fixed rate of rotation, and fixed small coefficient μ of sliding friction.

A too-naive interpration of grade school physics texts might leave one with the impression that the frictional work W during 1 revolution is

$$W = \mu \int_{\text{perimeter of gear}} F_{\text{normal}} ds \qquad \text{(Wrong)}$$

where s is the element of arclength and μ is the coefficient of friction. This would lead to the min-friction spur gear tooth curve being a logarithmic spiral $r = \exp(\theta_0 \pm \theta)$, and the min-friction bevel gear tooth curve being a loxodrome $r = 2 \arctan\left(\frac{1}{2}\exp(\theta_0 \pm \theta)\right)$. These curves are the spirals which cross radial geodesics at constant angle, namely 45°. But anyway, this is *wrong*; consider a contact consisting partly of rolling and partly of sliding, to see the problem[6].

The true frictional dissipation per unit time is $\mu V_{\text{sliding}} F_{\text{normal}}$ and thus the total dissipation during one revolution is

$$W = \mu \int_{\text{perimeter of gear}} \frac{V_{\text{sliding}} F_{\text{normal}}}{V_{\text{point of contact}}} ds. \tag{46}$$

Here it is understood that the integral is only to be extended over the part of the perimeter of the gear that is actually ever involved in sliding contact. If n active contacts happen, $n \geq 1$, then F_{normal} is reduced by a factor n, so that the total frictional work W, and our mathematical development, is unaffected.

Letting the curve of the gear in polar coordinates be $r(\theta)$ and assuming according to the fundamental law as usual that force is transmitted normally only, some easy trigonometry[7] gives us the elegant formula relating force vector to torque T

$$F_{\text{normal}} = T\sqrt{r^{-2} + r'^{-2}} \tag{47}$$

if r is the distance from the point of contact to the center of the gear. In the limiting case of a rack (whose curve is $y(x)$), this expression degenerates to

$$F_{\text{normal}} = F_{\text{linear}}\sqrt{1 + y'^{-2}}. \tag{48}$$

[6] The notion that involute gears would have greater efficiency if their racks were based on a characteristic angle of 45° instead of (the usual choices in engineering practice) 14.5° and 20° to the vertical, is, however, entirely correct.

[7] $T = rF_{\text{normal}} \cos\psi$ and see (EQ 2); for bevel gears $T = F_{\text{normal}} \cos\psi \sin r$ and see (EQ 3).

In the case of bevel gears, it is

$$F_{\text{normal}} = T\sqrt{r'^{-2} + (\sin r)^{-2}}. \tag{49}$$

The tooth-form which minimizes

$$W = \mu \int_{\theta_1}^{\theta_2} \frac{V_{\text{sliding}} F_{\text{normal}}}{V_{\text{point of contact}}} \frac{ds}{d\theta} d\theta \tag{50}$$

subject to specified values of $r(\theta_1)$, $r(\theta_2)$, $r'(\theta_1)$, $r'(\theta_2)$, is given by a certain differential equation which will be derived in §19.

16 Efficiency of gear trains

An interesting consequence of (EQ 46), (EQ 20), (EQ 47), and (EQ 44) is as follows. Assume $\mu \ll 1$. Also assume that the size G of the gear teeth is of the same order as s, which is of the same order as c, and that these are both much smaller than the pitch radius R of the gear.

Then the fraction of the transmitted power VF (where V is the pitch circle velocity, and F is the force transmitted along the pitch circles) between two spur gears that is lost to friction, is

$$(\frac{R_1}{R_2} + \frac{R_2}{R_1})G\kappa\mu. \tag{51}$$

where κ is a dimensionless "form factor" depending only on the shape of the tooth.

For bevel gears this would be

$$(\frac{\sin R_1}{\sin R_2} + \frac{\sin R_2}{\sin R_1})G\kappa\mu. \tag{52}$$

In both cases the quantity inside the parentheses is the reduction ratio plus its reciprocal. First, observe that smaller teeth result in greater efficiency. In the limit where the teeth vanish (gears reduce to smooth rolling wheels) there is zero frictional loss. Of course, the smaller the teeth, the lower their shear strength, and hence the smaller the allowable torque.

In any train of spur gears, each intermeshed with the preceding one, V, G, and the force F exerted along the pitch circles are all constant. Hence, if the reduction ratio R_1/R_2 is large, less power will be lost to friction if, instead of meshing gear 1 directly with gear 2, one instead places two idler gears, of radii R_3 and R_4, say, between gear 1 and gear 2, where $R_1/R_3 = R_3/R_4 = R_4/R_2 = X$. These two idler gears seem superfluous, but they will reduce the frictional dissipation from being proportional to $X^3 + X^{-3}$ to $3(X + X^{-1})$, a savings which, for 1000:1 reduction ($X = 10$) would be a factor of 33. Power savings will be obtained for any $X > \sqrt{2+\sqrt{3}} \approx 1.932$, i.e. if the overall reduction factor $X^3 > 7.21$. (Of course, this ignores friction in the bearings and only accounts for friction in the gears, hence X would have to be somewhat larger than this. In a train of 3 [instead of 4] gears, savings over simply meshing 2 gears occur when the overall reduction factor X^2 exceeds $(1+\sqrt{3}+\sqrt{2\sqrt{3}})^2/4 \approx 5.27$, although note the direction of rotation of the final gear will be reversed.)

In the more usual arrangement of gears, where (a smaller version of) gear 1 is meshed with gear 3, which is on the same shaft as gear 4, which meshes with gear 5, on the same shaft as gear 6, and this meshes with gear 2, and now only two sizes of gears are required and the radius ratio between meshed gears is again always X – in this case, the savings formula above would still hold since the smaller gears have smaller V but larger F in amounts which exactly cancel. Thus this usual arrangement has no advantage over the stupid one in terms of friction – its advantage is that smaller gears can be used.

The main point to come away with is the paradoxical (but true) conclusion of (EQ 51) that trains of gears are more efficient than a single gear pair, if the reduction ratio is large.

17 Uniform stress gear, assuming Hertzian contacts and linear friction

If the wear rate were some function of the normal force, the sliding speed, and the radii of curvature of the gear teeth at the point of contact, (and since most wear is actually a subsurface phenomenon depending on fatigue causing the creation of subsurface cracks, causing particles to break away... this is actually not a function of these things only, but in fact is a dynamic nonlocal function of the whole shape of the gear and its time history) one could try to make a gear profile on which the wear rate was constant.

Knowledge of tribology has not yet advanced to the point where we know what this function is. Hence, as a first approximation, we shall assume that repeated stress above a certain level, called the "surface fatigue limit," will cause wear, but stress below that level will never result in wear.

In the static linear theory of elasticity, arguments due to H. Hertz [6] [7] allow one to infer that if two cylinders of radii c_1,c_2 are in contact along a line and are sliding subject to a coefficient of friction μ, and also there is a normal force F, then the maximum stress in the cylinders will be proportional (the constant of proportionality depends on μ and elastic constants) to

$$\sqrt{F_{\text{normal}}}\sqrt{1/c_1 + 1/c_2}\,. \tag{53}$$

For contact of two coapical cones (this being relevant to bevel gears), this expression becomes

$$\sqrt{F_{\text{normal}}}\sqrt{\cot c_1 + \cot c_2}\,. \tag{54}$$

We imagine that the same relation holds when c_1 and c_2 are the radii of curvature of the gear teeth at the point of contact. Note that this holds, in the linear theory of elasticity with linear friction, *independent* of the sliding speed. We've assumed both cylinders/gears are made of the same material and the speeds involved are much smaller than the speeds of sound inside the materials so that we may treat the problem statically[8].

So if one believes that stress below some "fatigue limit" will never cause wear or fatigue, then we wish (to maximize allowable torque while never suffering from wear) for this to be constant. I call the resulting gears "uniform stress" but of course they are actually uniformizing the *maximum* stress.

We have previously got expressions for normal force (EQ 47) and radius of curvature (EQ 28) as a function of the form of the gear curve, and also (EQ 32) for the other gear's radius of curvature in terms of this gear curve. So we are now in a position to write down the function we want to uniformize. Specifically, let κ be a constant; then we want

$$\sqrt{r^{-2} + r'^{-2}}\left(\frac{1}{c_1} + \frac{1}{c_2}\right) = \kappa. \tag{55}$$

If c_1 and c_2 are expressed purely in terms of r, r', r'' by using previous formulae such as (EQ 2), (EQ 32), (EQ 33), (EQ 23), and (EQ 28), the result is

Theorem 8 *The tooth surfaces $r(\theta)$ of a gear of pitch radius R_1 with uniformized maximum stress, (assuming Hertzian contacts, constant coefficient of friction, the point of maximum stress is always at the contact, and that the conjugate gear has pitch radius R_2) and assuming small enough coefficient μ of friction that the transmitted torque is approximately constant (but not needing to assume that the law of friction is linear) obey the ordinary differential equation*

$$\kappa = \frac{(r'^4 + r^2 Q + r^3 r'' - r r'' Q + 2r'^2 Q)^2 R_1}{(r'^2 + r^2) r' r} / \tag{56}$$
$$\Big({R_1}^3 r^4 + 2{R_1}^3 r'^4 - 2r^3 r'^3 H - r^5 r' H - r r'^5 H + 2r^3 r'' Q R_1 - r^5 r'' R_1 - {R_1}^3 r^3 r''$$
$$+3{R_1}^3 r'^2 r^2 - r'^2 r^4 R_1 - 2r'^2 r^2 Q R_1 + r^3 r'' r'^2 R_1 + r'^4 Q R_1 - 3r'^4 r^2 R_1 - {R_1}^3 r r'' r'^2 - r^4 Q R_1 \Big)$$

where $1/H = 1/R_1 + 1/R_2$, κ is a constant, and

$$Q = \sqrt{{R_1}^2 r^2 + {R_1}^2 r'^2 - r^2 r'^2}. \tag{57}$$

Theorem 9 *For bevel gears, the differential equation is instead*

$$\sqrt{r'^{-2} + [\sin r]^{-2}}(\cot c_1 + \cot c_2) = \kappa \tag{58}$$

and where c_2 may be obtained from c_1 by (EQ 37), (EQ 22) and see also (EQ 3), (EQ 38), and

$$\tan c_1 = \frac{\sin r}{[\cos r - \psi'] \sin\psi} = \frac{\sqrt{r'^2 + \sin^2 r}}{\cos r - \psi'} = \frac{(\sin^2 r + r'^2)^{3/2}}{\cos r \sin^2 r - r'' \sin r + 2r'^2 \cos r}. \tag{59}$$

(I have not written down the full expression for (EQ 58) purely in terms of r, r', and r'', since it would be too long.)

I have been unable to solve the second order nonlinear ODE (EQ 55), or its bevel version (EQ 58), in closed form. But nothing stops one from solving it numerically.

Meanwhile, to compare ourselves with the Standard: Involute gears are quite non-optimal from the standpoint of stress. Just how non-optimal depends on how close to the bottom of the involute curve, contact is allowed to get. The radius of curvature at the bottom of an involute ($r \to r_0+$) approaches 0 proportionally to $\sqrt{r - r_0}$, causing the contact stress to grow toward ∞ like $(r - r_0)^{-1/4}$. Using very short involute segments far from the bottom of the involute would keep the contact stresses close to uniform, but the teeth of the resulting gear would be very small and hence would shear under large torques.

[8] If the two surfaces were instead *spheres* of radii c_1, c_2, the corresponding Hertzian stress would have been $\propto F_{\text{normal}}^{1/3}(1/c_1 + 1/c_2)^{2/3}$.

18 Gears with uniform temperature

Another process that can cause wear is for heat from friction to build up more quickly than it can dissipate. This is uncommon but it can happen in very high power gears. If that is the primary cause of wear, then the gears of theorem 10 which minimize heat production (frictional loss) would seem to be a good thing. But perhaps the problem is not the global heat production, but local heat production.

In this section we will imagine that the gears are made of a substance with very low thermal conductivity (such as glass?) and all the heat is carried away by the lubricant. However, we will assume that this (heat being carried away) can only happen at any particular point on a gear's surface *after* contact has ceased. The 'contact' will be Hertz-like. After contacxt is over, the removal of heat by the lubricant will be assumed to be sufficiently quick that all the heat is gone by the time contact reoccurs.

During contact, the rate of heat generation per unit area at a point is proportional to

$$\sqrt{F_{\text{normal}}}\sqrt{1/c_1+1/c_2}\,V_{\text{sliding}}\,\mu. \tag{60}$$

On the other hand, the duration a point is subjected to this rate of heat generation (i.e. endures contact) is proportional to

$$\sqrt{F_{\text{normal}}}\,(\sqrt{1/c_1+1/c_2}\,V_{\text{point of contact}})^{-1}. \tag{61}$$

Therefore, if no heat is dissipated until later, the total heat per unit area built up at a point during contact is proportional to the product of these two terms, that is, to

$$\frac{\mu F_{\text{normal}} V_{\text{sliding}}}{V_{\text{point of contact}}}. \tag{62}$$

This conclusion, we now see, would hold for a quite general kind of elastic contact, not necessarily Hertzian. It also holds mutatis mutandum for bevel gears. This leads us to the conclusion, in the limit of slow heat dissipation and high powered operation, that we want to uniformize the quantity in (EQ 62). Gears constructed in this way will be called "uniform temperature gears." Their tooth forms obey a differential equation which we will derive in §19.

19 ODEs for gears with minimal frictional losses or uniform temperature

The quantity $Z = \mu F_{\text{normal}} V_{\text{sliding}}/V_{\text{point of contact}}\, ds/d\theta$, may be expressed purely in terms of the curve $r(\theta)$ of gear 1, (and its derivatives r' and r'') and the pitch radii R_1 and R_2 of gears 1 and 2, by plugging in previously derived expressions for F_{normal} (EQ 47), V_{sliding} (EQ 20), $V_{\text{point of contact}}$ (EQ 45), the radius of curvature c (EQ 28), the distance s from the pitch point to the point of contact (EQ 23), and $ds/d\theta$ (where s now is arc length along the gear curve, sorry about the double use of the symbol "s") (EQ 25). The result (we are now considering the case of spur gears) is

$$Z \propto \frac{(R_2+R_1)(r^2-Q)}{(r^2+r'^2)R_2 r' r}\times \tag{63}$$

$$\frac{(2r^4+4r^2r'^2+r'^4-r^3r''-r^2Q+Qrr''-2Qr'^2)^2}{2r^4Q+{R_1}^2r^3r''-2r^6-3r^4r'^2-r^3Qr''+4r^2Qr'^2-3{R_1}^2r^2r'^2-r^3r''r'^2-{R_1}^2r^4-2{R_1}^2r'^4+{R_1}^2rr''r'^2+Qr'^4}$$

where Q was given in (EQ 57).

For "min friction gears," the integral of this quantity $d\theta$ from θ_1 to θ_2 is minimal over all choices $r(\theta)$ satisfying whatever boundary conditions are imposed at θ_1 and θ_2. One such choice of boundary conditions might be to specify $r(\theta_1)$, $r'(\theta_1)$, $r(\theta_2)$, and $r'(\theta_2)$. Consequently, the curve $r(\theta)$ of gear 2 must satisfy the Euler-Lagrange equations

$$(\frac{\partial Z}{\partial r})-\frac{d}{d\theta}(\frac{\partial Z}{\partial r'})+\frac{d^2}{d\theta^2}(\frac{\partial Z}{\partial r''})=0, \tag{64}$$

where Z is the quantity in (EQ 63). This is a 4th order nonlinear ordinary differential equation and almost certainly cannot be solved in closed form. However, there is no great obstacle to solving it numerically.

For "uniform temperature" gears, the quantity Z in (EQ 63), but with the factor $ds/d\theta$ omitted from the left hand side, i.e. divide the right hand side by $\sqrt{r'^2+r^2}$, is *constant* along the curve of the gear. (Except that, for portions of the gear curve at which there is an n-way multiple active contact, $n \geq 1$, the constant is n times larger, under the assumption that normal force is divided equally among the n active points of contact. As we have previously remarked in §15, the possibility of multiple contacts in no way affects the previously derived equation for min-friction gears.)

This statement is a 2nd order nonlinear ordinary differential equation. Appropriate boundary conditions for it might be to specify $r(\theta_1)$ and $r(\theta_2)$. Again, there is probably not a closed form solution and one must resort to numerical methods.

Notice that the value of the pitch radius R_1 of the other gear is immaterial (except inasmuch as it affects the value of the overall constant factor). Therefore, both uniform temperature gears and the gears with minimal frictional losses are "Don Juan."

Now let us consider the case of bevel gears.

The quantity $Z = \mu F_{\text{normal}} V_{\text{sliding}} / V_{\text{point of contact}} ds/d\theta$, may be expressed purely in terms of the curve $r(\theta)$ of gear 2, (and its derivatives r' and r'') and the pitch radii R_1 and R_2 of gears 1 and 2, by plugging in previously derived expressions for F_{normal} (EQ 49), V_{sliding} (EQ 19), $V_{\text{point of contact}}$ (EQ 42), the radius of curvature c (EQ 29), the distance s from the pitch point to the point of contact (EQ 21), and $ds/d\theta$ (where s now is arc length along the gear curve, sorry about that) (EQ 26).

We will not actually do write this out in full, since I have an aversion to expressions more than three lines long, but of course it is perfectly satisfatory to build it out of previous equations.

Theorem 10 *The spur gear tooth form with minimal frictional power loss under low friction conditions ($\mu \ll 1$, where μ is the coefficient of friction) is given by (EQ 64). For bevel gears the same ODE applies, except that the quantity Z in it is gotten as described two paragraphs ago.*

Technically, it should be remarked that we have not proven (yet) that gear tooth forms with minimal frictional losses *exist*, subject to the boundary conditions suggested near (EQ 64). But it is clear that one can make such existence statements subject to (for example) the restriction that we are only allowing 2-time differentiable monotonic functions $r(\theta)$ in the interval $\theta_1 < \theta < \theta_2$. This is because (1) from (EQ 20) and (EQ 47) we have a nonzero lower bound on the frictional loss among all tooth forms satisfying our boundary conditions at $\theta = \theta_i$, and (2) the set of allowed functions forms a compact set under suitable restrictions (3) the frictional loss (EQ 46) is a continuous functional under those restrictions (4) Any compact set over which is defined a continuous functional with a nonzero lower bound, has a minimum. Also, any minimum must obey the Euler-Lagrange equations (EQ 64). For the reader determined to find a gap in our reasoning, we remark that we have not shown that any solution of the Euler-Lagrange equations need be minimum. Merely some of them are.

Theorem 11 *The spur gear tooth form with uniform temperature as in §18 under low friction conditions ($\mu \ll 1$, where μ is the coefficient of friction) and assuming an elastic contact such as Hertzian, is given by $Z = \text{constant} \times ds/d\theta$ where Z is given in (EQ 63) and $ds/d\theta$ is in (EQ 25). For bevel gears the same ODE applies, except that the quantity Z in it is gotten as described four paragraphs ago and $ds/d\theta$ is from (EQ 26).*

20 A note on multiple contacts

In the derivation of theorems 8 and 11, we have ignored the possibility of multiple active points of contact. If there are $n \geq 1$ active points of contact, then under the assumption that the force is equally distributed between them, the shape formulas of these theorems will still be correct, *but*, in the tooth regions corresponding to n-way multiple contact, these shape formulas will need to be used *with different values of the constants* corresponding to the fact that the force is multiplied by $1/n$. This may result in more pieces in the piecewise analytic gear curve, but each piece will still be described by the same sort of equation.

21 Conclusion

Two shortcomings of the present paper are as follows.

Firstly, it is plausible that our uniform max-temperature and the uniform max-stress gears in fact *minimize* the maximum temperature and maximum stress (subject to the appropriate physical assumptions) over all possible gear curves satisfying the appropriate boundary conditions. But this is unproven.

Secondly, although we have found the mathematical form of the gear curves which optimize various properties (i.e. minimal frictional loss, uniform max-stress, etc.), any actual gear would have a curve whose pieces might individually be of the forms we have given, but precisely where best to select the boundaries between these pieces, and in such a way as to assure collision-free operation, is an issue we have not examined.

22 Acknowledgements

I would like to thank C.W. Gear for pointing out a big and stupid error in an early draft. Also David Waltz and Alan Clements of Harmonic Drive Technologies helped with footnote 2. I used the computer algebra system MAPLE.

References

[1] *American Institute of Physics Handbook*, McGraw-Hill. 1982 reissue.

[2] Earle Buckingham: *Analytical mechanics of gears*, McGraw-Hill 1949, Dover reprint 1963.

[3] S.J.Dokos: *Sliding friction under extreme pressure*, J.Appl.Mech. 13 (1946) A148-156,

[4] J.D.H. Donnay: *Spherical trigonometry*, Interscience 1945.

[5] D.W. Dudley (ed.): *Gear Handbook*, McGraw-Hill 1962.

[6] H. Hertz: *Über die Berührung fester Elastischer körper*, J. f. Math (Crelle) 92 (1882), also *Gessamelte Werke* vol. 1, Leipzig 1895 p.155,

[7] L.D.Landau and E.M.Lifshitz: *Theory of elasticity*, Pergamon, 3rd revised edition 1986.

[8] H.E.Merritt: *Gear Engineering*, Pitman 1971 (ISBN=0-273-42977-9).

[9] Derek J. de Solla Price: *An ancient Greek computer*, Scientific American (June 1959) 60-67.