APPENDIX A : Continued Fractions and Padé Approximations

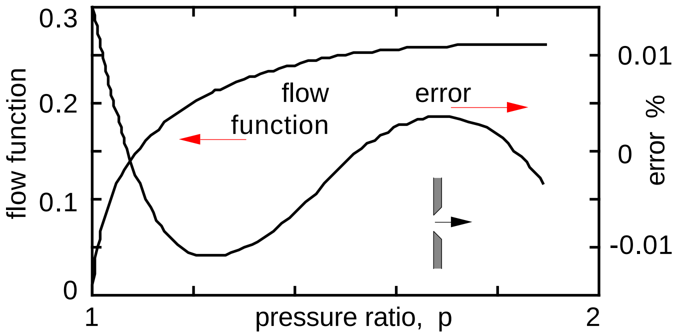
Truncating infinite power series in order to use the resulting polynomials as approximating functions is a well known technique. While some infinite series lend themselves to this, others do not - needing many terms and significant computing time to reach the desired accuracy. Some may indeed diverge. It is useful to realise that infinite series are not the only bases for numerical approximations, and it is for this reason that we now introduce the topic of ***continued fractions*** which often provide a **<u>much</u>** more efficient approximation basis than series. Continued fractions and power series are analogous to one another (compare (**iii**), (**iv**) below). ***Padé approximations*** are truncations of continued fractions, in the same way as polynomials are truncations of infinite power series.

The concept of a ***'convergent'*** arises from the study of continued fractions. Convergents find many applications - notably in finite elements - however their immediate relevance is as the basis of a procedure for selecting gear tooth numbers from a limited range to approximate irrational speed ratios. This procedure is carried out mechanistically however its grounding in convergents and continued fractions is reason enough to introduce these latter at this juncture.

We show, for interest, how a continued fraction may be derived from a power series, and the advantages of the resulting Padé expression over the corresponding polynomial as an approximation for a particular function. The reader is referred to Wall HS, **Continued Fractions**, van Nostrand 1948, for further details.

A typical example of the advantages of Padé approximations is the following.

The real-time dynamic simulation of the air-operated brakes of an ore train consisting of hundreds of cars requires thousands of iterations over the whole train. During each iteration the compressible **flow function** $\{ \mathbf{p}^{-2/\gamma} - \mathbf{p}^{-(1+1/\gamma)} \}$ has to be evaluated a number of times over each car to determine the flows through the various pipes and brake cylinder restrictions in the car - **p** being the variable **pressure ratio** across a restriction ($1 \le \mathbf{p} \le \mathbf{p}_{crit}$), and γ the constant ratio of specific heats for air. The flow function is plotted here.



The two exponentiations in the many functional evaluations took up such an inordinate amount of computing time that real-time simulation was jeopardised. To circumvent this the flow function was approximated by polynomials - three were required - these weren't particularly accurate but more significantly they gave rise to further computing convergence problems due to their piecewise nature.

A much superior approximation applicable over the whole **p**-range is of the form (**p** -1)(a +b.**p**) / (1 +c.**p**) in which a, b & c are carefully chosen constants. This is a Padé approximation and is indistinguishable from the the correct function at the scale of the plot - the accuracy of the approximation is for the most part within 0.01%, which is much superior to the polynomials.

In the context of gears, the foregoing expressions (**18**) (**22**) for tooth geometry factors are further examples where ***ratios*** of polynomials may be set up as Padé approximations to functions - in this case derived from finite element and photoelastic analyses.

A function of the single variable x may be written in the various forms :-

(**i**) $f(x) = a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \ldots$ a_i constant

(**ii**) $= x (a_1 + x (a_2 + x (a_3 + x (a_4 + \ldots))))$

(**iii**) $= c_1 x * (1 + c_2 x * (1 + c_3 x * (1 + c_4 x * (1 + \ldots))))$ c_i constant

(**iv**) $= b_1 x / (1 + b_2 x / (1 + b_3 x / (1 + b_4 x / (1 + \ldots))))$ b_i constant

(**v**) $$= \cfrac{b_1 x}{1 + \cfrac{b_2 x}{1 + \cfrac{b_3 x}{1 + \cfrac{b_4 x}{1 + \ldots}}}}$$

Form (**i**) is the standard infinite power series representation (neglecting any constant term) which must be truncated when used for computations - in which case it is often expressed more conveniently as (**ii**). Form (**iii**) is an alternate representation in which the c-coefficients may easily be derived from the a-coefficients; this form is included here only to show the similarity between the power series and the ***continued fraction*** (**iv**). Thus the multiplications which appear in the power series (**iii**) are replaced by divisions in the continued fraction (**iv**), which is often seen in expanded form (**v**).

If we set $b_4 = 0$ in (**v**), then the continued fraction is truncated after the third term - this is referred to as the third ***convergent*** :- $C_3 = b_1 x (1 + b_3 x) / (1 + (b_2 + b_3) x) \equiv A_3 / B_3$ and demonstrates the general form of the n'th convergent as a quotient between a numerator, A_n, and a denominator, B_n, both of which are polynomials. The convergent is an approximation of the original function - referred to as a Padé approximation.

A tabular procedure, involving only two simple recursive operations, may be used to derive the b-coefficients from the a-coefficients; that is to derive the continued fraction from the power series, since the latter approximations are widely available. The technique is exemplified below for a particular function, the natural logarithm, which is first expressed as an infinite power series of which six terms only (for example) are considered :-

$$f(x) = \ln(1+x) = x - x^2/2 + x^3/3 - x^4/4 + x^5/5 - x^6/6 + \text{etc.}$$

These a-coefficients are entered into the second row of the table below, the first row of which is a seed. Operations on succeeding rows are as follows.

An odd row is formed by dividing the preceding even row by its leading (leftmost) term - this term becoming the latest b-coefficient.
An even row is formed by subtracting terms in the preceding odd row from terms in the odd row before that, and moving the terms one column leftwards when entering them into the even row.

1			1	0	0	0	0	0
2			1	-1/2	1/3	-1/4	1/5	-1/6
3	$b_1 =$	1	1	-1/2	1/3	-1/4	1/5	-1/6
4			1/2	-1/3	1/4	-1/5	1/6	
5	$b_2 =$	1/2	1	-2/3	1/2	-2/5	1/3	
6			1/6	-1/6	3/20	-2/15		
7	$b_3 =$	1/6	1	-1	9/10	-4/5		
8			1/3	-2/5	2/5			
9	$b_4 =$	1/3	1	-6/5	6/5			
10			1/5	-3/10				
11	$b_5 =$	1/5	1	-3/2				
12			3/10					
13	$b_6 =$	3/10	1		etc			

Recurrence relations for the numerator (A_n) and denominator (B_n) of the n'th convergent are as follows :-

$A_0 = 0$	$B_0 = 1$
$A_1 = x b_1 = x$	$B_1 = 1$
$A_2 = A_1 + x b_2 A_0 = x$	$B_2 = B_1 + x b_2 B_0 = 1 + x/2$
$A_3 = A_2 + x b_3 A_1 = x + x^2/6$	$B_3 = B_2 + x b_3 B_1 = 1 + 2x/3$
$A_4 = A_3 + x b_4 A_2 = x + x^2/2$	$B_4 = B_3 + x b_4 B_2 = 1 + x + x^2/6$
$A_5 = A_4 + x b_5 A_3 = x + 7x^2/10 + x^3/30$	$B_5 = B_4 + x b_5 B_3 = 1 + 6x/5 + 3x^2/10$
$A_6 = A_5 + x b_6 A_4 = x + x^2 + 11 x^3/60$	$B_6 = B_5 + x b_6 B_4 = 1 + 3x/2 + 3x^2/5 + x^3/20$
etc	etc

The convergents (C_n) are finally expressed as ratios of polynomials. Approximations calculated therefrom, for some representative values of x, are tabulated :-

Trial variable	:	x	-1	1	2	5
Correct value	:	$\ln(1+x)$	$-\infty$	0.6931	1.0986	1.7918
Trunc. series	:	$x - x^2/2 + x^3/3 - x^4/4 + x^5/5 - x^6/6$	-2.1167	0.9500	15.73	3107.1
$C_1 = A_1 / B_1 = x$			-1	1	2	5

$C_2 = A_2 / B_2$	$= 2x / (2 + x)$	-2	0.6667	1	1.4286
$C_3 = A_3 / B_3$	$= x(6 + x) / (6 + 4x)$	-2.5	0.7	1.1429	2.1154
$C_4 = A_4 / B_4$	$= x(6 + 3x)/(6 + 6x + x^2)$	-3	0.6923	1.0909	1.7213
$C_5 = A_5 / B_5$	$= x(30 + 21x + x^2)/(30 + 36x + 9x^2)$	-3.3333	0.6933	1.1014	1.8391
$C_6 = A_6 / B_6$	$= x(60 + 60x + 11x^2)/(60 + 90x + 36x^2 + 3x^3)$	-3.6667	0.6931	1.0980	1.7787
etc.					

This exemplifies the superiority of an approximation based on a continued fraction over one based on a power series - for this particular function at least. It should be noted how successive “convergents” are alternately greater and less than the correct value (except for the infinite case), and converge upon it - hence their name.

The ability of continued fractions to identify simple ratio approximations, may be put to use for deriving gear trains for otherwise awkward ratios. The technique used is simpler than the above and is best shown by example. Suppose we require a ratio of 0.3711, correct to the fourth significant figure. We first express the fraction rationally - 3711 / 10000 - then factorise both numerator and denominator in an attempt to implement a compound train - $3 * 1237 / (2 * 5)^4$ - which gets us nowhere as 1237 is prime and far too large for a practical gear; we will have to use a simpler approximation. An extended division is therefore carried out below; initially dividing the ***numerator into the denominator*** and thereafter dividing the ***remainder into the previous divisor.*** The quotients thus found (ie. the b_n) form a continued fraction (of form different to (**v**)), whose convergents yield approximations to the original fraction.

```
3711 ) 10000 (  2 ≡ b1
       7422
       2578 ) 3711 (  1 ≡ b2
              2578
              1133 ) 2578 (  2 ≡ b3
                     2266
                      312 ) 1133 (  3 ≡ b4
                             936
                             197 ) 312 (  1 ≡ b5
                                   197
                                   115 ) 197 (  1 ≡ b6
                                         115
                                          82 ) 115 (  1 ≡ b7
                                                82
                                                33 ) 82 (  2 ≡ b8
                                                     66
                                                     16 ) 33 (  2 ≡ b9
                                                          32  etc
```

$$\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{3+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{2+\cfrac{1}{\text{etc}}}}}}}}}}$$

$$C_4 = \cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{3}}}} = 10 / 27$$

The convergents may be found as C_4 above - or by using the scheme below, in which C_0 is a seed, C_1 is the reciprocal of b_1, and thereafter the numerator and denominator follow from the same recurrence relation : $A_n = b_n A_{n-1} + A_{n-2}$; $B_n = b_n B_{n-1} + B_{n-2}$. The seventh convergent would evidently give the desired results -

Conv'gt order, n	0	1	2	3	4	5	6	7	8	9
Quotient, b_n	-	2	1	2	3	1	1	1	2	2
Numerator, A_n	0	1	1	3	10	13	23	36	95	226
Denominator, B_n	1	b_1=2	3	8	27	35	62	97	256	609
$C_n = A_n / B_n$	-	0.5	0.333	0.375	0.3704	0.37143	0.37097	0.37113	0.37109	0.37110
Proportional error	-	-3.5e-1	1.0e-1	-1.1e-2	2.0e-3	-8.9e-4	3.6e-4	-9.2e-5	1.7e-5	-4.4e-7

but it is impractical since 97 is prime and unavailable in our selection of change wheels. However the eighth is appropriate as $95/256 = 5 * 19/2^8$, and a compound train of 19/32 and 20/32 could be used.

Since all convergents are approximations to the same target value, their combinations may be more suitable than the convergents themselves - eg. $(C_8 + C_6)/2 = 59/159 = 0.3711$, and so on.

Admittedly, this technique has been largely superseded by cut-and-try methods on a computer.

APPENDIX B : Geometry of the Involute Gear Tooth

The equations which describe the profile of a rack-generated gear tooth, including both the involute and the fillet trochoid, are now derived. These are useful not only for power transmission gearing, but also for hydrostatic power transformation in gear pumps and so on, where fluid sealing and inter-tooth volume are important.

<u>Coordinates of a Point on the Rolling Rack</u> Any point such as Q, Fig B-1, is defined by coordinates (u, v) on the rack whose reference point, O, is offset by the profile shift, s, from the gear rolling circle at P. Thus for the fillet centre of Figure D :-

(i) $v_C = e - b$; $u_C = h + e \sec\alpha - v_C \tan\alpha$

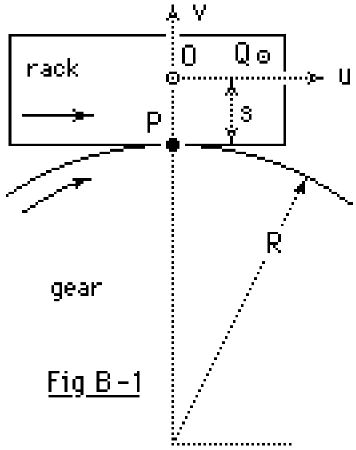


Fig B-1

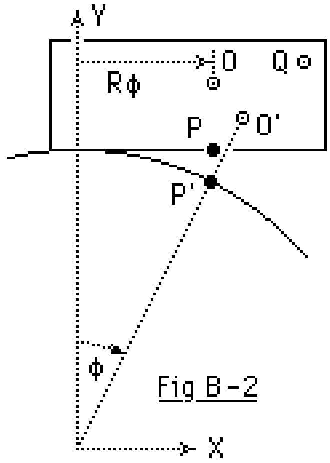


Fig B-2

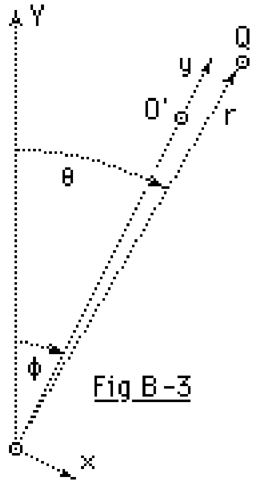


Fig B-3

It is required to determine the location of the Q <u>relative to the gear</u>, after a roll angle, φ, is undergone.

In Fig B-2, the coordinates (X, Y) of the point with respect to the gear centre are :

$X = R\varphi + u$; $R = z / 2$ z being the number of teeth on the gear, and

$Y = R + s + v$ which is constant

Transforming these to axes (x, y) fixed to the gear, Fig B-3, they become :-

(ii) $x = r \sin(\theta - \varphi) = X \cos\varphi - Y \sin\varphi = (R\varphi + u)\cos\varphi - (R + s + v)\sin\varphi$

$y = r \cos(\theta - \varphi) = X \sin\varphi + Y \cos\varphi = (R\varphi + u)\sin\varphi + (R + s + v)\cos\varphi$

So, the system parameters (α, a, b, e and h) having been laid down, and the gear parameters (R and s) specified, the coordinates of any point (u, v) may be found from (ii) for any roll angle, φ.

<u>The Involute Flank</u>

An involute is generated most simply by a point attached to a taught cord which unwinds from a base circle of radius $R_o = R \cos\alpha$, as shown in Fig B-4.

The polar coordinates of a point on the profile corresponding to the generation angle ψ, are :-

(iii) $r = R_o \sqrt{(1 + \psi^2)}$; $\theta = \gamma - \psi + \arctan\psi$

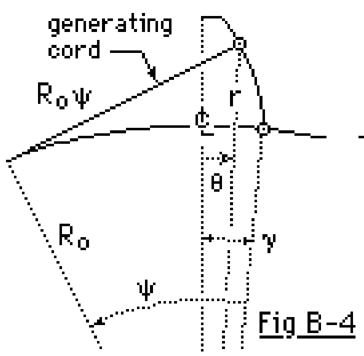


Fig B-4

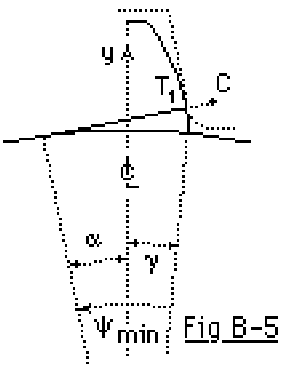


Fig B-5

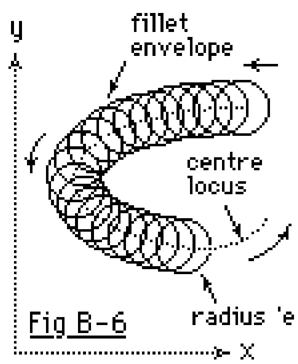


Fig B-6

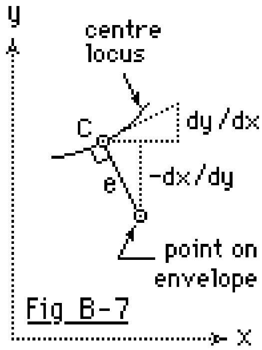


Fig B-7

With the rack in the symmetric position (φ = 0), the cord generator must be inclined at α to the horizontal, Fig B-5, since it is normal to the non-rotated rack at the contact point. It follows that the relation between the generation angle (ψ) and the rack roll angle (φ) is :-

(iv) $\psi = \varphi + \alpha + \gamma$

In order to determine the tooth constant half-angle, γ, the Cartesian coordinates of the contact point, calculated from (iii) in the symmetric position with $\psi = \alpha + \gamma$, are inserted into the equation for the rack flank derived from the geometry of Figure D as : $x = h + (R + s - y) \tan\alpha$. This yields the necessity :-

(v) $\gamma = (h + s \tan\alpha) / R + \mathrm{inv}\,\alpha$; $\mathrm{inv}\,\alpha \equiv \tan\alpha - \alpha$ . . . . the involute function

Checking for tip-pointing. The maximum generation angle, at the addendum circle, is, from (iii) :-

(vi) $\psi_{max} = \sqrt{ (r_a / R_o)^2 - 1 }$; $r_a = R + s + a$ . . . . the addendum radius

The teeth are pointed if the corresponding $\theta_{max} <= 0$, due in turn to the profile shift exceeding some critical s_{hi} which may be found, from (iii) and (v), to be :-

(vii) $s_{hi} = (R (\psi_{max} - \arctan\psi_{max} - \mathrm{inv}\,\alpha) - h) \cot\alpha$

noting that this must be solved by trial, since ψ_{max} is a function of s_{hi} here, via (vi).

On the other hand, the minimum generation angle, ψ_{min}, corresponds to the tangent point T_1 being tangent also to the involute, Fig B-5, provided the tooth is not undercut, that is provided that T_1 lies outside the base circle since the involute is not defined inside this circle. Comparing the ordinates of the fillet centre, C, with respect to the involute and to the rack :-

$y_C = R_o \cos\alpha + (R_o \psi_{min} + e) \sin\alpha = R + s - b + e$ from which

(viii) $\psi_{min} = \tan\alpha + (s + e (1 - \sin\alpha) - b) / R_o \sin\alpha$

and the corresponding lower limit for the trochoid, φ_{min}, follows from (iv).

Setting $\psi_{min} = 0$ in (viii) gives the critical profile shift, s_{lo} , below which undercutting occurs, thus :-

(ix) $s_{lo} = b - R \sin^2\alpha - e (1 - \sin\alpha)$

<u>The Trochoidal Fillet</u>

It may be appreciated from the sketches of tooth generation that the rack radius scours out the trochoidal tooth fillet, prior to the involute tooth flank being formed by the straight side of the rack. If C is the centre of the fillet circle, Fig B-6, then the coordinates of the corresponding point on the trochoidal fillet envelope are, from the geometry of Fig B-7 & (ii):

(x) $x = x_C + e\, dy/ds = \lambda (R \varphi + u_C) \cos\varphi - (R + \lambda (s + v_C)) \sin\varphi$ (u_C, v_C) from (i)

$y = y_C - e\, dx/ds = \lambda (R \varphi + u_C) \sin\varphi + (R + \lambda (s + v_C)) \cos\varphi$

where $\lambda = 1 + e / \sqrt{ (R \varphi + u_C)^2 + (s + v_C)^2 }$

The roll angle which defines the upper limit of the trochoid (φ_{max}) corresponds to the point T_2, Figure D, lying on the dedendum circle, radius $r_b = R + s - b$. Inserting coordinates of T_2 into (ii) leads to :-

(xi) $\varphi_{max} = - u_{T2} / R = - (h + e \sec\alpha + (b - e) \tan\alpha) / R$

<u>Undercutting</u>

If undercutting is indicated by (ix), then the intersection between involute and trochoid must be found by simultaneous solution of the intersection coordinates, via (iii) at ψ_{min} for the involute, and (x) at φ_{min} for the trochoid. Note that φ_{min} and ψ_{min} are <u>not</u> interrelated through (iv) in this instance.

<u>Calculation Sequence</u>

Given the system parameters (α, a, b, e, h); compute coordinates of C from (i).

Given the gear parameters (R, s); compute R_o, s_{lo} and s_{hi} via (ix) and (vii) respectively.

Set up the limits for the various profile segments :-

If $s < s_{lo}$ then the tooth is undercut; calculate φ_{min} from (x) and ψ_{min} from (iii) simultaneously
else determine ψ_{min} from (viii) and corresponding φ_{min} from (iv)

If $s >= s_{hi}$ then the tooth is pointed; calculate ψ_{max} from (iii) corresponding to $\theta = 0$
else determine ψ_{max} from (vi)

Ascertain φ_{max} from (xi)

Compute the various profile segments :-

The dedendum circle, radius r_b, from the tooth boundary at $\theta = \pi / z$, to the start of the trochoid; the trochoid from φ_{max} to φ_{min}; the involute from ψ_{min} to ψ_{max}; and finally, if not pointed, the addendum circle, radius r_a, to $\theta = 0$.