MODELLING DEPENDENCIES AND COUPLINGS IN THE DESIGN SPACE OF MESHING GEAR SETS

Mohammad Rajabalinejad
Assistant Professor, Faculty of Industrial Design Engineering, Delft University of Technology
Landbergstraat 15, Delft, the Netherlands
M.Rajabalinejad@tudelft.nl

ABSTRACT
This work presents a design methodology based on the combination of a set of compatibility equations for determining nominal tooth and cutter geometry and a set of tooth contact analysis equations for determining modified tooth surfaces and motion transmission laws. Both have been shown separately to lead to various optimisations, and some parametric subspaces of the designed gears are shown to be so weakly coupled that optimisations found individually may be superimposed, as shown in the case of the gear pair stiffness function, dynamical load factor, bending fatigue strength and pitting/scooring resistance optimisation. This is in contrast to traditional strengthening methods, such as profile shifting, which invariably produce much stronger couplings and thereby trade-offs.

KEY WORDS
Model development, gear design, profile modifications, gear meshing

1. Introduction
It has been shown recently [1-3] that the module-based calculus for gear geometry, stemming from practical manufacturing considerations, can only address a subset of the valid design space. It has also been shown that designs of significantly improved bending strength are to be found in the design space outside this subset, and it is therefore highly probable that other local optima besides bending strength exist and even, assuming that an agreement can be found on weighting factors of the defining criteria, a global optimum. Furthermore, dynamical loads are known to be highly dependent on the applied profile modifications as well, and recent developments of explicit solutions [4-5] for tooth contact analysis (TCA) and inverse tooth contact analysis (ITCA) make practical and straight-forward calculations possible. This paper discusses in some depth the new developments in calculus and performs a tentative fusion of the until-now-separate theories and procedures, building a theoretical framework for future optimisation studies. An important feature of the theory presented is that it can retain full compatibility and interchangeability with the standard 20° and 25° gear systems, even if the pressure angles and equivalent modules are different. This presents an important advance in gear geometric calculus, not only affecting our perception of the design space and possible embodiments, but, most importantly, paving the way towards cost-effective technological innovation.

2. Analytical Background
This work is based on two distinct theories, which are outlined as follows:

2.1 Involute compatibility equations
The compatibility of a rack-involute gear pair, as used typically in the study of gear generation, is broken down into compatibility of pitch and of various clearances (radial and backlash), as per Spitas [1]:

\[ t_g = t_g \cos \alpha_o \]  \hspace{2cm} (1)
\[ r_s \leq a_{IR} + \kappa_R + h_R \] \hspace{2cm} (2)
\[ r_f \leq a_{IR} + \kappa_R \] \hspace{2cm} (3)
\[ \frac{z}{2\pi} \alpha_o - \frac{a_{IR}}{t_g} \sin \alpha_o \leq \frac{1-c_{rg}}{2} \] \hspace{2cm} (4)

where \( t_g \) is the gear base pitch and \( \alpha_o \) is the rack half-angle (pressure angle). Such a rack-gear setup is shown in Fig. 1.
Further, the same analysis considers that two racks are equivalent, in the sense that they produce the same base pitch, if and only if:

$$t_R \cos \alpha_o = t'_R \cos \alpha'_o$$

(5)

where, in the above, the design parameters of the second rack are marked with an accent.

2.2 Tooth contact equations

Since recently, explicit solutions exist for the tooth contact analysis problem [4] and its inverse [5]. For the purposes of gear design the latter solution is quite useful, as it allows a desired function of transmission errors to singly and explicitly define the tooth profile needed to obtain such function, where previously one could only describe the modified tooth profile indirectly in terms of the tooling setup used to manufacture it. Refering to Fig. 2, this solution (in vector form) is as follows:

$$R_i \frac{dt}{d\theta_1} \times \left[ R_i f_i + \left( \frac{1}{f_{12}} \frac{d\theta_2}{d\theta_1} \right) (R_i f_i - a_{12}) \right] \cdot \hat{s}_3 = 0$$

(6)

This equation has the form:

$$U \left( \theta_1, \frac{d\theta_2}{d\theta_1}, r_2 \right) = 0$$

where $\theta_1$ is a solution parameter and $\frac{d\theta_2}{d\theta_1}$ is known from the problem definition. Eq. (6) is in all but the simplest cases implicit in terms of the unknown $r_1$, but may be solved reliably by standard numerical methods. Then the tooth profile may be expressed parametrically as $f_1(r_2)$, where:

$$f_2 = R_2^{-1} \left( R_1 f_1 - a_{12} \right)$$

(7)

$$r_2 = \| f_2 \|$$

(8)

and obviously the right-hand side of Eq. (7) is calculated as a function of $\theta_1, \theta_2(\theta_1)$ and $r_1(\theta_1)$.

Thus it is now possible to model the exact tooth surface contact problem, i.e. in a FEA environment, without linearisations (Litvin [6]) or other simplifying assumptions.

3. Analysis

3.1 Attributes and couplings of the design problem

Some important attributes relevant to gear design, together with ways to affect them, are shown in Table 1.

Stiffness by itself can affect the total angular displacement (transmission error) under a given load [7], as well as the real contact ratio [8]; depending of functional goals, it may be desirable to design for higher stiffness (accuracy), as shown in Table 1, or lower stiffness (self-correcting geometries). Most commonly, however, it is the stiffness function (the variation of stiffness with the angular position) that must be carefully designed, primarily to achieve a desired dynamical response [9].

In medium and high speed applications the dynamical load factor becomes an important attribute. It is a metric of the mesh self-excitations and it is generally desirable to minimise it, in order to reduce noise, vibrations and the fatigue loading of gears. It is strongly influenced by the stiffness function, and thereby by the contact ratio, any errors and profile modifications [9].

Bending fatigue resistance is dependent on both the loading conditions (i.e. dynamical load factor) and the tooth geometry and material. If it were possible to achieve the minimum load factor with the strongest tooth form, that would obviously result in the highest possible bending fatigue resistance; however the former
optimisation ideally requires slender, long addendum teeth (high contact ratio gears) [7, 10], while the latter ideally requires bulky high-pressure-angle teeth [7]. Clearly, the design parameters in this case are so strongly coupled, that the real optimum must be a compromise. On the other hand, strength optimisations exist for the tooth root that do not affect the working tooth form [1-3, 11-12], allowing for alternative routes to at least some local optima.

Like bending fatigue, pitting resistance also depends on the load factor, tooth form and material. Again, higher pressure angles are beneficial (higher radii of curvature at the contact point lower Hertzian pressure), but are not compatible with the high contact ratios usually required for minimum load factors. Unlike bending fatigue strength, the tooth root shape is irrelevant to pitting.

Designing for scoring resistance poses largely opposite requirements to bending and pitting resistance, because of its requirement to minimise sliding velocities. Thus design parameters are strongly coupled and compromises are inevitable.

Table 1. Methods to improve design attributes

<table>
<thead>
<tr>
<th>Design attributes</th>
<th>Methods to improve design</th>
<th>conventional</th>
<th>additional options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stiffness</td>
<td>Positive profile shift</td>
<td>Increase pressure angle (^2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increase cutter tip radius</td>
<td>Increase cutter half-angle as per Eq. (5) (^1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Profile modification</td>
<td>Reduce pressure angle (^2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long addendum</td>
<td>Reduce pressure angle (^2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative profile shift</td>
<td>Reduce base pitch</td>
<td></td>
</tr>
<tr>
<td>Dynamical load factor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bending fatigue strength</td>
<td>Positive profile shift</td>
<td>Increase cutter half-angle as per Eq. (5) (^1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increase pressure angle (^2)</td>
<td>Long addendum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increase cutter tip radius</td>
<td>Negative profile shift [13-14]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Profile modification</td>
<td>Profile modification</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) compatible with 20° standard, does not affect operating pressure angle. \(^2\) not compatible with 20° standard

3.2 Optimisation

Optimisation of multi-criteria multi-parametric design problems can be much more challenging than formulating an ad-hoc weighted function together with some mechanics-imposed constraints and dumping these into a numerical solver, because, if for no other reason, the multitude of criteria makes it practically difficult to properly weigh competing design goals; this is even more so in the context of a scientific paper, which must claim generality for its arguments and conclusions.

Thus we shall explore case-specific improvements and local solutions like weak Pareto optima [15]; this methodology is not only considered beneficial for its relative simplicity and clarity, but also because it is one that the industry more readily understands and can be expected to take an interest in and apply with little training.

The first step of this analysis is to reorganise the material from Table 1 as per its dependencies, as shown in Fig. 3 below.
4. Case study and discussion

As has been mentioned before, formulating a complete optimisation methodology is outside the scope of this paper. The ground work will be laid down, however, in the following case study, to test and discuss several important underlying aspects.

Consider a pair of typical 20° spur gears having \( z_1 = 25 \) and \( z_2 = 50 \) teeth respectively and tooth thickness coefficients \( c_{s1} = c_{s2} = 0.5 \). Addendum \( c_{ki} \), dedendum \( c_{fi} \) and cutter tip radius \( c_{i} \) coefficients as standard. Non-dimesionalising with respect to the module (equivalently: assuming module equal to unity), the nominal centre distance will be \( a_{12} = 37.5 \). This shall be henceforth called embodiment #0. The design parameters are reported in Table 2. The base radii are \( r_{g1} = 11.746 \) and \( r_{g2} = 23.492 \). The contact ratio is \( \varepsilon = 1.683 \). Obviously, gear material, fabrication method, surface treatment, speed, loading and lubrication are some of the many factors contributing to the attributes listed in Table 1. For simplicity, let us consider such a design context where the safety factors for bending fatigue, pitting and scoring are all equal to unity. This has the advantage of further non-dimesionalising the following treatment.

![Fig. 3. Dependencies in the design space](image)

This illustration immediately shows that all the design attributes in discussion here are strongly coupled through profile shifting and changing of the pressure angle. The former is exceedingly popular in design practice, while the technological importance of the latter is evident in the existence of the 14.5° and 25° involute systems. It is therefore unavoidable that using any of the above two methods to strengthen one attribute (i.e. bending strength) weakens another (i.e. scoring resistance). The controversy is even stronger in the context of pitting or bending optimisations, where obtaining higher contact ratio is at odds with obtaining a stronger tooth (root form and flank curvature). Balancing these compromises is the typical object of conventional optimisations [14].

However, it can be observed that other methods such as increasing the cutter half-angle as per Eq. (5), applying targeted profile modifications as per Eqs. (6)-(8) and increasing the cutter tip radius are free from undesired couplings and allow for independent optimisations along the Pareto frontier [2-3, 9], hence Pareto-efficient solutions. This allows for less reliance on profile shifting and/ or pressure angle changes, which also introduce incompatibilities between mating gears and require further (sometimes extensive) supporting modifications.

In effect, by making use of the new methods presented in section 2, new dimensions are added to the design space that allow to by-pass compromising constraints. This will be shown in the following case study.

|  |  |  |  |  |  |
|---|---|---|---|---|
| 1 | 25 | 1.000 | 0.000 | 0.500 | 0.300 | 1.250 |
| 2 | 50 | 1.000 | 0.000 | 0.500 | 0.300 | 1.250 |

Table 2. Embodiment #0: a nominal design

Let us first consider a formal optimisation for bending stress, i.e. as per [13]. From the stress tables in that work we obtain the optimum design shown in Table 3. This will serve as embodiment #1.
Obviously, the combined change in tooth proportions results in a different contact ratio \( \varepsilon = 1.699 \). The HPSTC are located at radii \( r_{h1} = 12.319 \) and \( r_{h2} = 25.432 \). By applying unitary loads at the HPSTC, the maximum non-dimensional tensile stress can be calculated for each gear by means of FEA, as seen in Fig. 4. To best account for usual tooth widths, plane strain conditions are assumed; each model is meshed using 6-node triangular isoparametric finite elements distributed as per a variable density self-correcting algorithm, so that maximum computational error is kept below 2%. For comparison, the corresponding standard tooth designs are analysed as well. At the same positions the radii of curvature are also calculated (in relation to pitting resistance), as are the sliding velocities (in relation to scoring resistance), as per the following equations:

\[
\begin{align*}
p & \propto \frac{1}{\sqrt{\rho_i}} \quad (9) \\
\rho_i^* & = \frac{1}{\sqrt{\rho_i^2 + \frac{1}{\rho_{ij}^2}}} \\
\text{where} \\
\rho_i & = \sqrt{r_{h1}^2 - r_{hi}^2} \\
\rho_{ij} & = \frac{r_{h1} + r_{h2}}{\tan(\pi/2 - \alpha_o)} - \rho_i
\end{align*}
\]

Regarding the scoring risk, this is calculated at the tip of each gear in terms of the scoring factor (Lingaiah [16]):

\[
PVT \propto (\rho_i - r_i \sin \alpha_i) \sigma_{hi} \quad (10)
\]

where

\[
\sigma_{hi} \propto \frac{1}{\sqrt{\rho_i r_i (a_{ho} \sin \alpha_e - \rho_i)}}
\]

and the radii of curvature \( \rho_i \) are in this case calculated at the tooth tips.

![Fig. 4. FEA of gear tooth. HPSTC loading, base fixed.](image)

The results are as follows for embodiment #0:

\[
\sigma_1 = 3.10, \quad \sigma_2 = 2.63, \quad \sigma_{crit} = 3.10 \quad \text{(critical root bending stress)}, \quad \rho_1 = 4.51, \quad \rho_{12} = 8.32, \quad \rho^*_1 = 3.96, \quad \rho_2 = 8.93, \quad \rho_{23} = 3.86, \quad \rho^*_2 = 3.54, \quad p_{crit} = 0.51 \quad \text{(critical surface stress based on } \rho^*_1), \quad \sigma_{H1} = 0.07, \quad PVT_1 = 0.17, \quad \sigma_{H2} = 0.10, \quad PVT_2 = 0.27, \quad PVT_{crit} = 0.27 \quad \text{(critical scoring factor)}
\]

For embodiment #1:

\[
\sigma_1 = 2.54, \quad \sigma_2 = 2.52, \quad \sigma_{crit} = 2.54, \quad \rho_1 = 3.71, \quad \rho_{12} = 9.12, \quad \rho^*_1 = 3.44, \quad \rho_2 = 9.74, \quad \rho_{23} = 3.09, \quad \rho^*_2 = 2.95, \quad p_{crit} = 0.58 \quad \text{(critical surface stress based on } \rho^*_2), \quad \sigma_{H1} = 0.07, \quad PVT_1 = 0.11, \quad \sigma_{H2} = 0.14, \quad PVT_2 = 0.46, \quad PVT_{crit} = 0.46 \quad \text{(critical scoring factor)}
\]

Thus it can be seen that, while embodiment #1 has increased bending resistance, it has adversely affected pitting and scoring resistance.

An alternative embodiment #2 can be found by applying the calculus in section 2.1. Such an embodiment may be realised by cutting the 25-tooth pinion with a 25° equivalent cutter (new module \( m_1 = 1.037 \)), so that it may substitute the original pinion directly in the pair; the
mating gear need not be modified. It has already been shown by Spitas [1] that the working profile and mesh kinematics remain unaffected; therefore the path of contact, contact ratio, pitting and scuffing conditions will be the same as in the case of embodiment #0. The different root geometry does affect the bending strength, however, and is therefore calculated, as before, by means of FEA. The result is $\sigma_i = 2.83$, therefore $\sigma_{crit} = 2.83$ (critical root bending stress).

Table 4. Embodiment #2 as per [1]

<table>
<thead>
<tr>
<th>$i$</th>
<th>$z_i$</th>
<th>$m_i$</th>
<th>$x_i$</th>
<th>$c_{ii}$</th>
<th>$c_{ki}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>1.037</td>
<td>0.000</td>
<td>0.380</td>
<td>0.520</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>1.000</td>
<td>0.000</td>
<td>0.500</td>
<td>1.000</td>
</tr>
</tbody>
</table>

These results, together with their projected effect on the corresponding safety factors, with embodiment #0 as reference, are shown in Table 5.

Table 5. Attribute comparison of different embodiments

<table>
<thead>
<tr>
<th>Embodiment #</th>
<th>Bending fatigue</th>
<th>Pitting</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\sigma_{crit} = 3.10$</td>
<td>$p_{crit} = 0.51$</td>
<td>$PVTT_{crit} = 0.27$</td>
</tr>
<tr>
<td></td>
<td>$S_\sigma = 1.00$</td>
<td>$S_p = 1.00$</td>
<td>$S_{PVTT} = 1.00$</td>
</tr>
<tr>
<td>1</td>
<td>$\sigma_{crit} = 2.54$</td>
<td>$p_{crit} = 0.58$</td>
<td>$PVTT_{crit} = 0.46$</td>
</tr>
<tr>
<td></td>
<td>$S_\sigma = 1.22$</td>
<td>$S_p = 0.88$</td>
<td>$S_{PVTT} = 0.59$</td>
</tr>
<tr>
<td>2</td>
<td>$\sigma_{crit} = 2.83$</td>
<td>$p_{crit} = 0.51$</td>
<td>$PVTT_{crit} = 0.27$</td>
</tr>
<tr>
<td></td>
<td>$S_\sigma = 1.10$</td>
<td>$S = 1.00$</td>
<td>$S_{PVTT} = 1.00$</td>
</tr>
</tbody>
</table>

It can be seen that embodiment #2 does not improve the bending resistance as much as embodiment #1, but it does not compromise pitting and scoring resistance either. In this sense, it is more Pareto-efficient.

Since embodiment #2 has not altered the working profiles or their kinematics, other optimisations of the standard design (embodiment #0) may be superimposed with no undesired couplings or trade-offs. The next embodiment will attempt to manipulate the load factor.

Embodiment #3:

In [9] a study was presented to minimise the load factor by means of profile modifications in the form of tip relief. The recommendations were made based on dynamical simulations and took into account the existence of index errors of various magnitudes and of an optimal ‘flat region’ where sensitivity to the modification parameters is minimised. Tooth geometry can be studied and prescribed as per the equations of section 2.2. Assuming the same dynamical system and conditions as in those simulations, and a maximum index error of i.e. 15μrad, the load factor for embodiment #0 at a speed of about 950rpm is calculated to be 2.4. Optimally, a tip relief of 15μrad will reduce the load factor to 1.85. The corresponding safety factor for bending will thereby increase by a factor of $\frac{2.4}{1.85} = 1.30$ and the factors for pitting and scoring by $\sqrt{1.30} = 1.14$. This, like embodiment #2, is also Pareto-efficient.

But why is Pareto efficiency so important? Surely optimisations exist, where it is quite acceptable to trade-off one positive attribute for another; this is especially common in the technical world and in engineering design. At the same time, however, we must recognise that there is much we do not know. I.e. the formula for pitting may be indicating that a given pinion is reasonably safe, to a degree that we may be tempted to make it less so to strengthen its root, if we feel that that is too weak; but, in fact, any gains are at best limited to +20% (i.e. see Table 5). Given our limited capacity to predict the exact operating conditions in most applications, as well as the statistical nature of material strength, or the interdependency between failure modes, it therefore does not always make sense to trade strength off: here the philosophical advantage of Pareto optimisation becomes apparent. This is a philosophy already embraced by many industries (in a large part because the technical industry is the fusion ground between economics and technology) and it is therefore reasonable to expect that the methodology leading to embodiments #2-3 and beyond can be picked up and applied efficiently.
Not least, the possibility to simply superimpose different design optima, as outlined in this work, is also appealing to any practical engineering design process: gains of previous optimisations are thereby preserved.

It should be noted for the sake of completeness that several optimisation alternatives other than that used in embodiment #1 exist, both in the scientific literature [17-18] [14] and in industrial practice (DIN [19], AGMA [20] standards, for a comparison see Spitas [21]). However, all are based on the use of profile shifting, which unavoidably weakens the positively shifted gear(s) in terms of pitting and scoring. Therefore, while the actual numerical predictions would vary between scenarios, the overall picture of non-Pareto-efficient solutions, as in embodiment #1, would persist.

5. Conclusion

This work presented a design methodology based on the combination of a set of compatibility equations for determining nominal tooth and cutter geometry and a set of tooth contact analysis equations for determining modified tooth surfaces and motion transmission laws. These were shown to lead to Pareto-efficient designs with advantages over those obtained by traditional strengthening methods, such as profile shifting, which invariably produce strong couplings of the design attributes.

References

[5] Spitas C., Costopoulos T., Spitas V., Direct analytical solution of the inverse gear tooth contact analysis problem, Inverse Problems in Science and Engineering, 16(2) (2008), 171-186
[8] Spitas C., Costopoulos T., Spitas V., Calculation of transmission errors, actual path of contact and actual contact ratio of non-conjugate gears, VDI- Berichte, 1665 (2002), 981-994