GENERALIZED PARTICLE SWARM ALGORITHM FOR HCR GEARING GEOMETRY OPTIMIZATION

Siniša Kuzmanović1, Miroslav Vereš2, Milan Rackov1

1Faculty of Technical Sciences, University of Novi Sad, Dositeja Obradovića 6, 21 000 Novi Sad, Serbia, s.kuzmanovic@open.telekom.rs, racmil@uns.ac.rs
2Faculty of Mechanical Engineering, Slovak University of Technology, Nám. slobody 17, 812 31, Bratislava, Slovakia, miroslav.veres@stuba.sk

Key words: HCR gearing, contact ratio coefficient, pitting, optimization.

Abstract Temperature scuffing evidenced by damage to teeth flanks of gears is one of the most important problems needing to be solved in the process of gearing design and calculation. According to current valid standards, such calculations can be resolved with a high level of reliability for all the usual gearing types. However, suitable calculations for HCR gears have not been adequately researched to date. It has been identified that in HCR gears some different process of scuffing formation occurs during the gear’s operation. In this article, the authors describe a new method for finding optimal solutions for $h_{1s}$, $h_{2s}$, and $x_1$, using a Generalized Particle Swarm Optimization Algorithm.

1 INTRODUCTION

Standard gear transmissions are used with a normal contact ratio, i.e. a contact ratio between 1 and 2. Load sharing between the teeth is possible, but there is a time during meshing when one pair of teeth takes the entire load. To ensure a smooth and continuous operation, a minimum contact ratio $\varepsilon_\alpha$ of 1,4 is preferred. Contact ratios for conventional gearing are generally in the range 1,4 to 1,6. High contact ratio (HCR) gears are non-standard involute gears with a transverse contact ratio higher than, or equal to two ($\varepsilon_\alpha \geq 2$), which can be achieved by increasing the gearing addendum tooth height $h_{1s}$ to more than 1, and/or by reducing the normal pressure angle $\alpha_n$ below the standard value of 20°. The design of HCR gears is complicated due to the fact there is a greater possibility of mesh interference occurring, than there is in standard involute profiles. Also here there is a higher risk from a too small thickness of tooth tip, and significantly less favorable slip ratio values, which can cause scuffing damage to the flanks.

2 HIGH CONTACT RATIO (HCR) OF INVOLUTE GEAR PROFILE

The correct working of the gear will be assured if the value of the contact path is higher than that of the base pitch (Fig.1). The ratio between the length of contact $g_{se}$ and pitch on the base cylinder $p_{bt}$ is given by the following formula:

$$
\varepsilon_\alpha = \frac{g_{se}}{p_{bt}} = \frac{\sqrt{r_2^2 - r_1^2} + \sqrt{r_2^2 - r_3^2} - a_s \sin \alpha_{st}}{p_{bt}}
$$

(1)

It is called the transverse contact ratio. Contact ratios for conventional gearing are generally in the range 1,4 to 1,6; so the number of tooth engagements is either one or two. One special kind of involute, non-standard gearing is called high contact ratio (HCR) gearing, where the contact ratio is higher, and there are always at least two pairs of teeth in contact ($\varepsilon_\alpha \geq 2$).
The advantage of HCR gearing is a higher load capacity (load distribution is shared between more pairs of teeth at the same time) and a lower relative gearing noise level something which can be significantly reduced by using an integer HCR factor $\varepsilon_\alpha$. The geometry of involute HCR gearing and the general distribution of applied force at important points of the HCR gear tooth flank are shown in Fig. 1. The biggest applied force in HCR gearing (between points BB’ and DD’) can be considered to be about 50%, when two pairs of teeth are in contact. Consequently, the size of the applied force is decreased when three pairs of teeth are in contact. So, the value of involute gearing $2/3F$ is decreased to $1/3F$, and the value $1/3F$ is decreased to $1/6F$. When $\varepsilon_\alpha>2$, three pairs of teeth are in contact along the line of action, this is triple tooth contact. Triple tooth contact has a generally higher load capacity and lower gearing noise level. Another favorable property of this HCR gearing is the increased resistance to fatigue damage. In high precision and heavily loaded spur gears, the effect of gear error is negligible, therefore the periodic variation of tooth stiffness is the principal cause of noise and vibration. High contact ratio gears could be used to reduce the variation of tooth stiffness.

3 HCR GEARING AND THE POSSIBILITIES FOR OPTIMIZING ITS GEOMETRICAL PARAMETERS

According to the results of different measurements of gear pairs, noise reduction proved to be best when HCR gearing with the value of contact ratio $\varepsilon_\alpha=2$ is used. The increase in contact ratio can be implemented in two ways: by decreasing the pressure angle, and by increasing tooth height, that is: $\varepsilon_\alpha=f(h_a, \alpha)$. The most favorable solution is obtained by increasing addendum height, however there are a lot of geometrical and manufacturing constraints which limit the possibilities of increasing the contact ratio. Equation (1) shows that $\varepsilon_\alpha=f(g_a, p_h)$. The tooth pitch at the base circle of standard involute gearing is equal to the base pitch on HCR gearing, and it is considered as constant. This means that achieving the greatest value of the contact ratio $\varepsilon_\alpha$ has to be obtained through the greatest possible increase in the length of the line of action $g_a$. The length of the contact line $g_a$ is calculated in equation (2):

$$g_a = \sqrt{r_1^2 - r_{a1}^2} + \sqrt{r_2^2 - r_{a2}^2} - a_w \sin \alpha_w$$

(2)

where the tip diameters of the pinion and gear wheel are as follows:

$$r_{a1} = r_1 + (h_{a1} + x_1 \cdot m_a)$$
$$r_{a2} = r_2 + (h_{a2} + x_2 \cdot m_a)$$

(3)

From equations (2) and (3), it is clear that the length of the contact line $g_a$ is directly dependent on the addendum height $h_{a1}$, $h_{a2}$ and the addendum modification factors $x_1$, $x_2$. Optimization of the geometric parameters of HCR gearing can be based on objective function in order to get the maximum value of contact ratio $\varepsilon_\alpha$ for a given centre distance $a_w$. The main optimization parameter at this level can be the addendum height of the teeth $h_{a1}$, $h_{a2}$, and addendum modification factors $x_1$ and $x_2$. For a given distance between the centers of the wheels, $x_c=x_1+x_2$ can be expressed as a function of $a_w$.

Addendum heights $h_{a1}$ and $h_{a2}$ can be derived from equations:

$$h_{a1} = h_{a1}^* \cdot m_a$$
$$h_{a2} = h_{a2}^* \cdot m_a$$

(4)

where $h_{a1}^*$ and $h_{a2}^*$ are addendum heights for modules equal to one. That means that the contact ratio is the objective function of both addendum heights and the addendum modification factor of the pinion.
As for all conventional involute gearing, the thickness of HCR teeth on the tip circle is given by the risk of abruption of the tip of the tooth. As mentioned earlier, the size of the unit addendum height can be expressed:

\[ h_{a1}^* = \frac{r_{a1} \cdot \sin^2 \alpha_{a1}}{m_n \cdot \cos \alpha_{a1}} + x_1, \quad h_{a2}^* = \frac{r_{a2} \cdot \sin^2 \alpha_{a2}}{m_n \cdot \cos \alpha_{a2}} + x_2 \]  

To avoid meshing interference, the following conditions must be satisfied:

\[ h_{a1}^* \leq \frac{1}{m_n} \left[ \frac{a_{a1} \sin \alpha_{a1} - r_n - \tan \alpha_{a1} \sin \alpha_{a2}}{\sin \alpha_{a2} \sin \alpha_{a1} - \sin \alpha_{a2} \sin \alpha_{a1}} \right] - x_2 \]  
\[ h_{a2}^* \leq \frac{1}{m_n} \left[ \frac{a_{a2} \sin \alpha_{a2} - r_n - \tan \alpha_{a2} \sin \alpha_{a1}}{\sin \alpha_{a1} \sin \alpha_{a2} - \sin \alpha_{a1} \sin \alpha_{a2}} \right] - x_1 \]  

where:

\[ \cos \alpha_{a1,2} = \frac{r_{a1,2}}{r_{a1,2}}, \quad \alpha_{a1,2} = -\arccos \frac{r_{a1,2}}{r_{a1,2}}, \quad \alpha_{a1,2} = \arccos \left( \frac{r_{a2,1}}{r_{a2,1} + h_{a2,1} + x_{2,1}} \right) \]

Changing addendum height, \( h_{a1}^* \) will certainly influence the total thickness of the tooth on the tip circle. Greater tooth height, as well as a positive addendum modification factor, may cause a reduction in the thickness of the tooth on the tip circle to below the permissible value. Tooth tip thickness can be expressed:

\[ s_{a1} = d_{a1} \left( \frac{p_i + 4 x_i m_n \tan \alpha_{a1}}{2 d_1} + \inv \alpha_{a1} - \inv \alpha_{a1} \right) \]

As for all conventional involute gearing, the thickness of HCR teeth on the tip circle is given by the condition \( s_{a1} \geq \frac{0.25 - 0.4}{m_n} \). For case-hardened teeth this value should be \( s_{a1} \geq 0.4 m_n \) due to the lower risk of abruption of the tip of the tooth. As mentioned earlier, the size of the unit addendum height \( a_{a1} = f(h_{a1}^*) \). Thus, for case-hardened teeth, the inequation becomes:

\[ 0.4 m_n \leq d_{a1} \left( \frac{p_i + 4 x_i m_n \tan \alpha_{a1}}{2 d_1} + \inv \alpha_{a1} - \inv \alpha_{a1} \right) \]

and after arranging:

\[ h_{a1}^* \leq \frac{p_i + 4 x_i m_n \tan \alpha_{a1}}{2 d_1} - \frac{d_1}{2 m_n} - x_2 \]  

A similar pattern is valid for the unit addendum height \( h_{a2}^* \).

\[ h_{a2}^* \leq \frac{p_i + 4 x_i m_n \tan \alpha_{a2}}{2 d_2} - \frac{d_2}{2 m_n} - x_2 \]

### 4 GENERALIZED PARTICLE SWARM OPTIMIZATION ALGORITHM

The optimization method used in this paper requires the use of the MATLAB program and is a relatively new optimizing method called the Generalized Particle Swarm Optimization Algorithm (GPSO). This GPSO algorithm was implemented in MATLAB. The algorithm was implemented as a function which allows the user to specify the values of all the adjustable parameters and to define many available algorithm options. Successful optimizers are often inspired by natural processes and phenomena. Indeed, the natural world is extraordinarily complex, but it provides us with remarkably elegant and robust solutions to even the most difficult problems. The field of global optimization has benefitted greatly from these nature-inspired techniques, such as Genetic Algorithms (GAs), Simulated Annealing (SA), Ant Colony Optimization (ACO) and others. Among these search strategies, the Particle Swarm Optimization (PSO) algorithm is relatively novel, yet well studied, and is a proven
optimizer based on the social behavior of animals moving in large groups (particularly birds). Compared to other evolutionary techniques, PSO has only a few adjustable parameters, and is computationally inexpensive and very easy to implement. PSO uses a set of particles called a swarm to investigate the search space. Each particle is described by its position (x) and velocity (v). The position of each particle is a potential solution, and the best position that each particle achieves during the entire optimization process is memorized (p). The swarm as a whole memorizes the best position ever achieved by any of its particles (g). The position and the velocity of each particle in the k-th iteration are updated as

\[ x[k+1] = w \cdot v[k] + c_p \cdot r_p[k] \cdot (p[k] - x[k]) + c_g \cdot r_g[k] \cdot (g[k] - x[k]) \]  \hspace{1cm} (12)

\[ x[k+1] = x[k] + v[k+1] \]

Acceleration factors \( c_p \) and \( c_g \) control the relative impact of the personal (local) and common (global) knowledge on the movement of each particle. Inertia factor \( w \), keeps the swarm together and prevents it from diversifying excessively and therefore diminishing PSO into a purely random search. Random numbers \( r_p \) and \( r_g \) are mutually independent and uniformly distributed over the range [0, 1]. Most theoretical analyses reported in the literature are conducted on the basis of the second-order PSO model

\[ x[k+1] = (1 + w - c_p \cdot r_p[k] - c_g \cdot r_g[k]) \cdot x[k] + w \cdot x[k-1] = c_p \cdot r_p[k] \cdot p[k] + c_g \cdot r_g[k] \cdot g[k] \]  \hspace{1cm} (13)

which is equivalent to the model described in (12). The current paper addresses a generalization of PSO that is based on the general linear discrete second-order model that is well known from control theory. The algorithm itself was proposed by the authors in [2], where it was called Generalized PSO (GPiSO).

The idea of this algorithm is in fact to address PSO in a new and conceptually different fashion, i.e., to consider each particle within the swarm as a second-order linear stochastic system with two inputs and one output. The output of such a system is the current position of the particle (x), while its inputs are personal and global best positions (p and g, respectively). The stability and response properties of such a system can be directly related to its performance as an optimizer, i.e., its exploitative and explorative properties. Thus, one can overcome the inherent flaw of the PSO algorithm, which is its inability to independently control various aspects of the search, such as stability, oscillation frequency and the impact of personal and global knowledge. Equation (13) can be interpreted as a difference equation describing the motion of a stochastic, second-order, discrete-time, linear system with two external inputs. In general, a second-order discrete-time system can be modeled by the recurrent relation

\[ x[k+1] + a_1 \cdot x[k] + a_2 \cdot x[k-1] = b_1 \cdot p[k] + b_2 \cdot g[k] \]  \hspace{1cm} (14)

where some or all of the parameters \( a_1, a_2, b_1 \) and \( b_2 \) are stochastic variables with appropriate, possibly time-varying probability distributions. Several restrictions should be imposed on these parameters in order to make (14) a successful optimizer. Firstly, the system should be stable and its stability margins should grow during the optimization process. The swarm should explore the search space first, and the particles should initially be allowed to move more freely. As the search process approaches its final stages, the optimizer should exploit the good solutions that were previously found, and the effort of the swarm should be concentrated in the vicinity of the known solutions. Secondly, the response of the system to the perturbed initial conditions should be oscillatory so that the particles can overshoot or fly over their respective attractor points. Further, if both external inputs approach the same limit value as \( k \) grows, the particle position x should also converge to this limit. Finally, in the early stages of the search, the system should primarily be governed by the cognitive input \( p \), allowing particles to behave more independently in these stages. In later stages, the social input \( g \) should be dominant because the swarm should become more centralized, and the global knowledge of the swarm as a whole should dominate the local knowledge of each individual particle. All of these requirements are formulated in a sequence using notions from control theory. The characteristic polynomial of (14) is

\[ f(z) = z^2 + a_1 \cdot z + a_0 \].

The dynamics of each particle is primarily defined by the roots of this polynomial, which are also known as the eigenvalues of the system. The system is stable if, and only if, the modulus \( \rho \) of the eigenvalues is less than 1. In order for the system to be able to oscillate, the roots of the characteristic polynomial must be negative real numbers. The argument \( \phi \) of the eigenvalues determines the discrete frequency of the characteristic oscillations of the system. Argument values close to \( \pi \) result in more frequent oscillations. The requirement that particle positions tend to the global best position when personal best and global best are equal is the same as the requirement that the system (14) has unit gain when both of its inputs are equal. This is equivalent to

\[ 1 + a_1 + a_0 = b_1 + b_2 \]  \hspace{1cm} (15)
It is also clear that an increase in \( b_p \) favors the cognitive component of the search, while an increase in \( b_g \) favors the social component. All of the requirements can easily be satisfied if system (14) is rewritten in the following canonical form, often used in control theory [17]:

\[
x[k + 1] = \rho x[k] + \rho^2 x[k - 1] + (1 - 2\rho) (c \cdot p[k] + (1 - c) \cdot g[k])
\]

In (56), \( \rho \) is the eigenvalues module, and \( \zeta \) is the cosine of their arguments. Parameter \( c \) is introduced to replace both \( b_p \) and \( b_g \). Clearly, requirement (15) is satisfied by (16). The primary idea of the GPSO algorithm is to use (16), instead of (12) or (13), in optimizer implementation. The parameters in this equation allow a more direct and independent control of the various aspects of the search procedure. For simplicity, both attractor points (i.e., global and personal best) are assumed to be equal to zero. Note that lower values of parameter \( \rho \) lead to faster convergence, while higher values result in less stable motion and slower convergence. Thus, higher values of \( \rho \) would enable the swarm to cover a wider portion of the search space. Lower values of \( \rho \) are beneficial in the later stages of the search, when faster convergence is preferable. Parameter \( \zeta \) determines the way particles oscillate over attractor points. For \( \zeta \) equals 1, a particle would approach its attractor in a non-oscillatory fashion. For \( \zeta \) equals -1, a particle would oscillate erratically from one side of the attractor to another. By adjusting \( \zeta \) to a value between -1 and 1, a desired behavior can be obtained. It is known that proper parameter selection is crucial for the performance of PSO. The relationships between adjustable factors within classical PSO and GPSO are straightforward, with

\[
\rho = \sqrt{\frac{w}{c}} \quad \zeta = \frac{1 - w - cp \cdot rp - cg \cdot rg}{2\sqrt{w}} \quad c = \frac{cp \cdot rp + cg \cdot rg}{cp \cdot rp - cg \cdot rg}
\]

The inertia factor \( w \) is the “glue” of the swarm; its responsibility is to keep the particles together and to prevent them from wandering too far away. Since its introduction by Shi and Eberhart [3], it has been noted that it is beneficial for the performance of the search for this to be gradually decreased in value. A common choice is to select a value close to 0.9 at the beginning of the search and a value close to 0.4 at the end. Based on (17), it would be reasonable to designate \( \rho \) as decreasing from about 0.95 to about 0.6. Regarding the acceleration factors, \( cp=cg=2 \) is the choice used in the original variant of the PSO algorithm. Other commonly used schemes include \( cp=cg=0.5 \) as well as \( cp=2.8 \) and \( cg=1.3 \). It is generally noted that the choice of acceleration factors is not critical, but they should be chosen so that \( cp+cg<4 \). In fact, if the last condition is violated, PSO does not converge, regardless of the inertia value. Based on recommendations stated earlier, two parameter adjustment schemes are proposed, with \( \rho \) linearly decreasing from 0.95 to 0.6 and \( c \) linearly decreasing from 0.8 to 0.2. The proper selection of \( \zeta \) proved to be the most difficult task, as this factor has no direct analogy to any of the parameters of the classical PSO. The \( \zeta \) parameter equals the cosine of the particle eigenvalues in the dynamic model. Because this is a discrete system, the eigenvalue argument equals the discrete circular characteristic frequency of particle motion. Thus, \( \zeta \) is directly related to the ability of particles to oscillate around attractor points (i.e., global and personal best) and therefore has a crucial impact on the exploitative abilities of the algorithm. If \( \zeta \) equals 1, this would prevent particles from flying over the attractor points; however, if \( \zeta \) is close to -1, this would result in the desultory movement of the particles. In both cases, particles cannot explore the vicinity of the attractor points. Based on thorough empirical analysis and numerous numerical experiments conducted by the authors, it is clear that the most appropriate and robust choice is to select \( \zeta \) uniformly distributed across a wider range of values. Based on many analyses of PSO and its parameters, two parameter adjusting schemes, GPSO1 and GPSO2, were proposed in [3]. In both of them, a linear decrease in \( \rho \) from about 0.95 to about 0.6, and in \( c \) from about 0.8 to approximately 0.2. is recommended. In the first scheme (GPSO1), \( \zeta \) was adopted as a stochastic parameter with uniform distribution ranging from [-0.9, 0.2], while in the second scheme (GPSO2), \( \zeta \) was uniformly distributed in the range [-0.9, 0.6].

5 OPTIMIZATION OF HCR GEOMETRY

For optimization of HCR geometry, the following data is used: number of pinion teeth \( z_1=21 \), number of gear wheel teeth \( z_2=51 \), module \( m=4 \) mm, modified centre distance \( a_c=144 \) mm, pressure angle \( \alpha_p=20^\circ \) and helix angle \( \beta=0^\circ \). The goal is to achieve \( \varepsilon_r=2 \). With this value of contact ratio it is expected that a reduction in the vibration and gear noise will be achieved. In order to achieve high contact ratio, addendum height is made higher in order to obtain a longer contact path. So, there are several variable parameters:

- addendum height divided by module on the pinion teeth \( h_{a1} \in <1, 1.5> \)
- addendum height divided by module on the gear wheel teeth \( h_{a2} \in <1, 1.5> \)
rack shift factor of the pinion – \( x_1 \in \langle -1, 1 \rangle \), and their values should be obtained in order to achieve a contact ratio equal to two. 

\[
\varepsilon_a = \frac{z_1}{2 \pi} \left[ \tan \alpha_{at1} - \tan \alpha_{at1} - t \cdot \tan \alpha_{at2} + t \cdot \tan \alpha_{at2} \right] = 2
\]

(18)

Using the GPSO algorithm, the solution of the HCR value is obtained in a very short time, less than one second. This solution is very accurate, and it achieves \( 10^{-15} \) accuracy. The results obtained using this method are presented below.

Given data: \( z_1 = 21, z_2 = 51, \)\( mn = 4 \text{ mm}, aw = 144 \text{ mm}, \)\( \alpha = 0.349065850398866 \text{ rad}, \beta = 0 \)

Calculated optimal values for 100 iterations:

- \( h_{a1} = 1.184548908194918 \)
- \( h_{a2} = 1.313253162560083 \)
- \( x_1 = 0.174169864574093 \)
- Main function: \( \epsilon_{alpha} = 2.000000052991015 \)

Calculated optimal values for 500 iterations:

- \( h_{a1} = 1.17779276410901 \)
- \( h_{a2} = 1.31951635807650 \)
- \( x_1 = 0.17860257195356 \)
- Main function: \( \epsilon_{alpha} = 2.00000000000000 \)

6 CONCLUSION

According to the results obtained and their analysis, it is possible to state that the optimization of HCR gearing geometry from the point of view of minimizing temperature scuffing damage is an important task in transmission design. Optimizing the results to fully meet the requirements stated in the objective function, when the results meet the defined conditions, demonstrates that the whole process to find the optimal vector \( \left( h_{a1}^*, h_{a2}^*, x_1 \right) \) is limited by the small number of suitable combinations of optimizing parameters. Finding the global minimum value of the objective function is usually a complicated problem [20]. However, swarm-based methods in general, and PSO in particular, provide us with powerful and robust tools for tackling the global optimization problems encountered in science and engineering. PSO however has some disadvantages. The key disadvantage is its inability to independently control various aspects of the search, such as the stability, oscillation frequency and the impact of personal and global knowledge. The new algorithm, the Generalized PSO (GPSO), which is described and analyzed in detail in above [2], overcomes this aforementioned flaw.

The method presented here of finding optimal solutions for \( h_{a1}^*, h_{a2}^* \) and \( x_1 \) uses the Generalized Particle Swarm Optimization Algorithm and MATLAB as a program for optimization. This GPS optimization is a very fast and reliable method. Using this method provides an infinite number of very precise solutions to the transcendental equation, and they all offer a solution for contact ratio \( \varepsilon_a = 2 \).

Acknowledgement

This contribution has been developed as a result of project 1/0277/12 subtask, supported by the Slovak VEGA grant agency.

REFERENCES

[1] Vereš, M., Kuzmanović, S., Rackov, M.: Experimental Research of HCR Gearing From Pitting Damage Point of View. Proceedings of 7th International Symposium about Mechanical and Industrial Engineering, KOD 2012, ISBN 978-86-7892-399-9, 317-320.

[2] Kanović, Ž, Rapajić, M.R., Jeličić, Z.D. Generalized Particle Swarm Optimization Algorithm - Theoretical and Empirical Analysis with Application in Fault Detection. Applied Mathematics and Computation 217, (2011) 10175-10186.

[3] Schutte, J.C., Groenwold, A.A., A Study of Global Optimization Using Particle Swarms, Journal of Global Optimization 31, (2005) 93-108.