Chapter 2: Review of Literature

The modern engineering use of polygonal shafts was limited, due to complex profile geometry that needed dedicated grinding machines, compared to the ease to produce alternatives such as splined and keyed shafts until the development of numerically controlled machining. In 1939, a kinematically controlled grinding machine (Maximov, 2005) for the production of triangular profiles with filleted corners, called K- profiles (named after the company Krause–Vienna, which introduced this type of joint (Huang et al, 2010)), was developed. One of the early studies on these profiles was conducted in Germany in 1958 by E. Filemon (1959). In the 1960s, epicycloidal profiles with three and four lobes were first made and the DIN standard 32711 for three lobe (P3G profile) and 32712 for four lobe (P4C profile) polygonal profiles were published in 1979 (Maximov,2005). Polygonal shafts still had limited use due to difficulty in manufacturing until the advent of modern manufacturing techniques and processing. Using computer numerically controlled manufacturing, the polygonal shafts can be precisely and economically manufactured and is being used in industries due to its advantages.

2.1) Analytical Stress Analysis

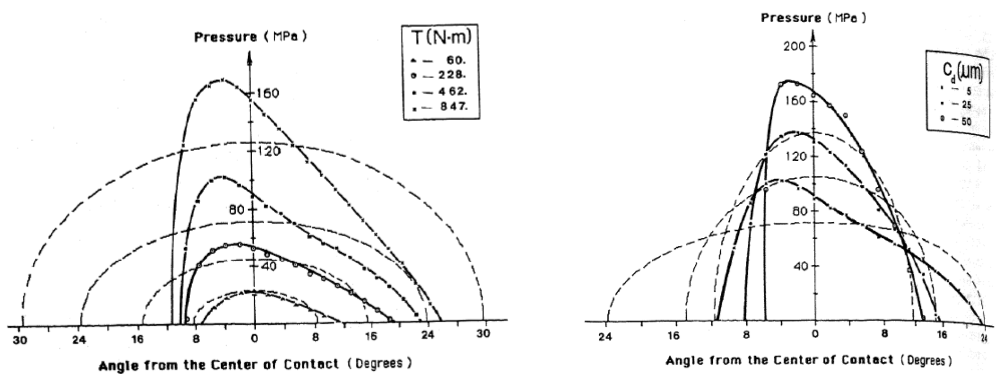
Stress analyses of polygonal profiles were developed by various writers such as Orlov, Leroy, Viseur, Musyl and Manhurim (Citarella & Gerbino, 2005) and were based on very strong approximations that are not able to clarify the real stress and strain state. The aim of the analysis was to find the critical stress in the hub since hub was supposed to expand under torsion and fail. The procedure attempted to simplify the geometry of the polygon connection by analogous mechanical models. For example, Musyl used circular segments for profile approximation. This approach of Musyl is referred by the polygonal connection manufacturers (Taylor, 1987). The current DIN standard is also developed on the works of Musyl and provides an approximation of contact pressure and to compute nominal stress state for static torsional loads (Standard DIN 32711 & 32712, 2009-03). Later, Ziaei calculated the stress state using conformal mapping for a shaft subjected to torsion (as cited in Grossman, 2006). All these analyses did not consider the more realistic loading condition of torsional bending. In the absence of an accurate analytical solution of the polygonal shaft, numerical methods have been used to determine the loading stress.

2.2) Numerical Stress Analysis

Numerical studies on the polygonal shaft hub connection were performed using the finite element method and the boundary element method. The numerical analysis conducted in polygonal connections has been explained in two sections; viz. 2D stress analysis and 3D stress analysis.

2.2.1) 2D Stress Analysis

Due to the complex conformal nature of the contact between the shaft and the hub in a polygonal connection, significant work in the contact pressure distribution in polygonal shaft and hub connections was not conducted until the advent of numerical methods. The contact being conformal, does not fall under Hertz contact theory and has to depend upon numerical solution. With the popularity of numerical methods, more realistic stress and strain analyses in polygonal shafts were developed. The initial numerical studies by Braschel et al. (as cited in Grossman, 2007) and Czyzewski and Odman (1988) were restricted to the two dimensional analysis that omitted the possibility of stress concentration at the end of the hub contact. Czyzewski and Odman (1988) did an analysis on contact stress and deformation on trilobe polygonal connections, which was the first published solution to the contact problem in trilobe connection (Czyzewski & Odman, 1988). The study consisted of an in-plane torsional loading without considering the friction and bending moment and showed the effect of torque and clearance on the contact pressure. The result found that the contact pressure is approximately triangular in distribution and does not resemble Hertz contact theorem, being a highly conformal contact. Czyzewski and Odman (1988) recommended the value of maximum shear stress for determining the load capacity of the connection. The effect of torsional loading and diametrical clearance on contact pressure distribution found from the study are shown in Figure 9.



(a) (b)

Figure 9: Effect of (a) Torsional loading and (b) Diametrical Clearance on pressure distribution in zone of contact (Czyzewski & Odman, 1988).

(Permission to reproduce image is in Appendix E)

Huang et al. (2010) studied the effect of friction on contact stress in a trilobe connection. Similar to the studies conducted by Czyzewski and Odman (1988), Huang et al. (2010) found the triangular distribution of contact stress and found that friction decreases the value of the normal contact stress, although the distribution and rule of contact stress were unaffected. They also found an increase in contact area and decrease in the normal contact force as the eccentricity increased. The analysis of shearing stress in polygonal shaft and hub was conducted by Lü and Liu (2011) using plane stress analysis for concentrated load at three points in trilobe polygonal shaft and hub connection and found higher shear stress in hub than the shaft.

The comparison of trilobe polygonal shafts to the involute splines for shafts with mean diameter of the polygonal shaft equal to the pitch diameter of the involute splines were performed by Kahn-Jetter, Hundertmark and Wright (2000) and found tensile stress in the polygonal connection to be significantly lower than in the spline due to the lack of stress concentration. This showed a positive trait for using polygonal connections. These studies were only two dimensional and did not depict the realistic behavior of the connection and the need for three dimensional study was inevitable.

2.2.2) 3D Stress Analysis

Mechnik conducted a three dimensional analysis, mainly on three lobe polygonal profiles, and found the peak stress at the edge of the hub and shaft connection. This study was mainly concentrated on the effect of the shape of the polygon on stress values and suggested increased profile eccentricity for better load carrying capacity (as cited in Grossman, 2007). Citarella and Gerbino (2001) used boundary element analysis to determine the state of stress and strain, which came in close conformance to the finite element results of Mechnik, but with lesser computation than FEA. The study of Gottlichar was a more realistic three dimensional study, since it accounted for the bending load, which is inevitable in most of the transmission. This showed the shaft as the most endangered part due to fretting corrosion exposure (as cited in Grossman, 2007). Grossman (2007) used the torsional bending load for both three and four lobe polygonal connections and looked at the most optimized shape in terms of loading capacity. He found that the global profile shape determined the fatigue strength of the connection more than the manufacturing precision.

2.3) Conformal Contact Problems

The highly conformal geometry of a polygonal shaft and hub excludes the necessary assumptions to be analyzed by Hertzian contact theory. A contact is said to be conforming if the surfaces of the two bodies 'fit' exactly or even closely together without deformation as in polygonal shafts. The nature of contact is a closely conformal concentrated contact of convex and concave surfaces. Although asperities occur in these type of conformal contacts, such asperities are so small relative to the geometry that they can be neglected during the analysis (Grossman, 2007).

Due to the presence of the appreciable fraction of the circumference of the shaft and the hub in contact, the elastic half space theory cannot be applied. Also, the effect of friction cannot be neglected and the body tends to be outside the scope of Hertzian contact theory. In the limiting case where the clearance between the shaft and the hub is large, the concept of non-conformal contact can be used to approximate the solution.

To find the contact stress for Non-Hertzian elastic bodies, the analytical form needs initial separation to be described in simple quadratic form, which is not possible in the epitrochoidal curve. Hence, computational contact mechanics is applied to analyze the system.

2.4) Computational Contact Mechanics

The need of nonlinear analysis arises due to the dynamic stresses that occur because of changing contact areas (topology nonlinearity) for which there is no analytical solution. The stiffness of the structure is a function of displacement and is no longer constant and the solution is nonlinear.

The contact and target elements are to be defined where the contact approaches the target because the two bodies do not interpenetrate. To avoid the penetration, ANSYS® uses various methods of contact formulation. The pure penalty method uses a virtual spring that provides stiffness so that the body does not penetrate. Another contact formulation, the Normal Lagrange adds an extra degree of freedom, contact pressure, so as to satisfy the variational inequality. Combining the advantages of both these methods, Augmented Lagrangian formulation can be used, which defines the normal push back force as a combination of the normal stiffness and Lagrangian contact pressure as shown in equation (38):

$$F_n = k_n x_n + \lambda \qquad (38)$$

where, F_n is the normal push back force, k_n is the stiffness and x_n is the displacement of virtual spring from Penalty method, λ is the contact pressure from Lagrangian formulation.

This method is less sensitive than the pure penalty method and is preferred for frictional contact stress problems (ANSYS INC, 2010). The way by which the computational software addresses non linearity is by dividing a load step into a number of time steps (sub steps), which have several iterations for each linearized equilibrium condition. If the solution do not converge in these sub steps, the bisection method is enabled to define new sub steps. Computational software such as ANSYS® uses the Newton Raphson method to solve the problem, which is a series of linear approximations with corrections. It shall be clearer from Figure 10, which shows the Newton Raphson method for a single sub step. The iterations continue for force and displacement convergence for a given criteria. The reason for iteration is the nonlinearity. In a nonlinear analysis, as in the contact problem, the stiffness is a function of displacement and is not constant. The equation is given as:

$$[K(U_i)]\{U_i\}=\{F_i\} \qquad (39)$$

where, U_i and F_i are the i^{th} displacement and force, where i is the current equilibrium iteration.

Initially the displacement is at U_o and the external force is F_o and the point on the response curve is P_o. The software increases the external force by $F_o+\Delta F$ so that using the Jacobian or tangent stiffness, $K(U_o)$, the displacement is calculated as:

$$[K(U_o)]\{\Delta U\}=\{\Delta F\} \tag{40}$$

The displacement is increased to $U_1=U_o + \Delta U$ and the point is (U_1, $F_o+\Delta F$) in P' as shown in Figure 10. Then, the value of displacement is substituted back in equation (39) to get the actual force F_1. The difference between the force, $F_o+ \Delta F$ and F_1 is called residual. If the residual force is less than the criterion, the subset is said to be converged. Otherwise, the next iteration is carried on with the value of U_1, F_1 as the new initial point.

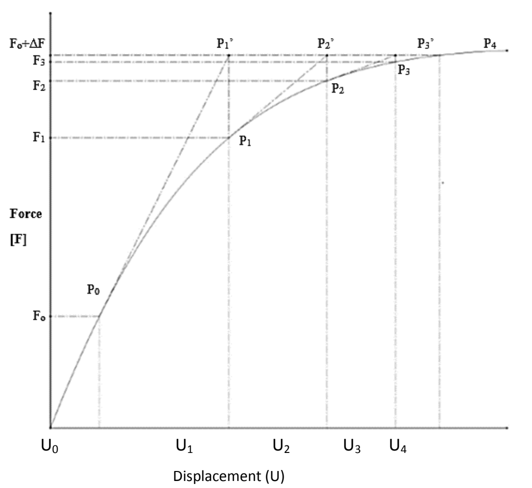


Figure 10: Newton Raphson method for solving nonlinear problems