INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING, VOL. 31, 525–545 (1991)

A COMBINED SURFACE INTEGRAL AND FINITE ELEMENT SOLUTION FOR A THREE-DIMENSIONAL CONTACT PROBLEM

SANDEEP VIJAYAKAR

Advanced Numerical Solutions, 2085 Pine Grove Lane, Columbus, Ohio 43232, U.S.A.

SUMMARY

A method is described to determine contact stresses and deformation using a combination of the finite element method and a surface integral form of the Bousinesq solution. Numerical examples of contacting hypoid gears are presented.

INTRODUCTION

In earlier studies[1, 2, 10, 11] a pure finite element approach was used to obtain compliance terms relating traction at one location to the normal displacement at another location. It became apparent that, in order to obtain sufficient resolution in the contact area, the size of the finite element model would have to be inordinately large. A finite element mesh that is locally refined around the contact region cannot be used when the contact zone travels over the surfaces of the two bodies.

Other researchers working in the tribology area[3, 7, 9] have obtained compliance relationships by integrating the Green's function for a point load on the surface of a half space (the Bousinesq solution) over the areas of individual cells demarcated on the contact zone. This method works well as long as the extent of the contacting bodies is much larger than the dimensions of the contact zone, and the contact zone is far enough from the other surface boundaries so that the two contacting bodies may be treated as elastic half spaces. These conditions are, however, not satisfied by the bodies being considered here.

The approach that is described here is based on the assumption that, beyond a certain distance away from the contact zone, the finite element model predicts deformations well. The elastic half space model is accurate in predicting *relative* displacements of points near the contact zone. Under these assumptions, it is possible to make predictions of surface displacements that make use of the advantages of both the finite element method and the surface integral approach.

This method is related to asymptotic matching methods that are commonly used to solve singular perturbation problems. Schwartz and Harper[8] have used such an asymptotic matching method to determine the relative approach of two rigid cylinders pressed against an elastic cylinder in plane strain. This method also bears resemblance to some classical approaches.[12, 13]

In order to combine the surface integral solution with the finite element solution, a reference or 'matching' interface embedded in the contacting body is used. This matching surface is far enough removed from the principal point of contact so that the finite element prediction of displacements along this surface is accurate enough. At the same time, it is close enough to the principal point of

0029–5981/91/030525–21$10.50

Received 12 February 1990
Revised 22 May 1990

contact so that the effect of the finite extent of the body does not significantly affect the relative displacements of points on this surface with respect to points in the region of contact.

This paper describes the analysis of the contact problem, and some numerical results obtained for a set of hypoid gears. Some contact stress contour diagrams, contact patterns and transmission error predictions are shown.

PRELIMINARIES

Contact analysis is carried out in several steps. The first step is to locate a set of 'primary contact points', which are points at which the individual contacting surfaces are closest to each other before the application of load. The next step is to calculate the principal normal curvatures of the individual surfaces, and the unit vectors along the tangents to these surfaces in the directions of maximum and minimum principal normal curvature. Then the relative curvature between the two surfaces is determined, and the unit vectors along the directions of maximum and minimum principal normal relative curvatures are calculated. The next step is to estimate the size of the contact zone. This is done by using Hertz's model for contact. Using this rough estimate of the contact zone, a grid is laid out around each principal contact point (Figure 1). Then cross compliance terms between the various grid points are calculated using a combination of a surface integral form of the Bousinesq and the finite element model of the contacting bodies. Finally, load distributions and rigid body movements are calculated using the algorithm based on the Simplex method.[10]

(*a*) *Locating the primary contact points*

The surface of each of the contacting bodies is broken into separate parts, each of which is likely to make contact with a part of the other body. In the case of gears, each such part is the surface of just one side of one gear tooth. Each of these is searched for a 'principal contact point'. This point is where the two surface patches

$$\mathbf{r}_1(s_1, t_1) \text{ and } \mathbf{r}_2(s_2, t_2) \quad -1 \leqslant t_1, t_2 \leqslant +1, 0 \leqslant s_1 \leqslant n_1, 0 \leqslant s_2 \leqslant n_2 \tag{1}$$

come the closest to each other.

In other words, the distance $\|\mathbf{r}_1(s_1, t_1) - \mathbf{r}_2(s_2, t_2)\|$ between the two surfaces is extremized with respect to the four variables s_1, t_1, s_2, t_2. This search can be carried out by the method of

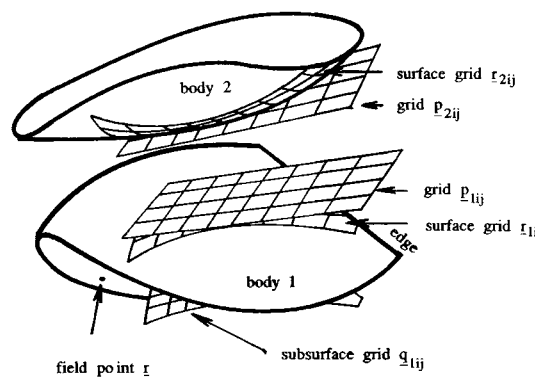


Figure 1. The contacting bodies and the computational grids

steepest descent for the four variables. A special purpose finite element was used to numerically represent these surfaces by C^1 continuous (smooth) functions. But even so, using the method of steepest descent was ridden with convergence problems. It was too sensitive to numerical error. Another method is to exhaustively search the set spanned by these four variables. This method is too time consuming. A combination of the two methods was found to work the best. The space (s_1, t_1) was broken into a grid of points (s_{1i}, t_{1j}) which corresponds to a grid $\mathbf{r}_{1ij} = \mathbf{r}_1(s_{1i}, t_{1j})$ drawn on the surface $\mathbf{r}_1$. For each of these grid points, an attempt was made to locate $\mathbf{r}_{2ij} = \mathbf{r}_2(s_{2i}, t_{2j})$ that minimized

$$\|\mathbf{r}_1(s_1, t_1) - \mathbf{r}_2(s_2, t_2)\|$$

with respect to the variables s_2, t_2. This extremization is equivalent to solving the following pair of non-linear equations:

$$\begin{aligned} &(\mathbf{r}_{1ij} - \mathbf{r}_2(s_{2i}, t_{2j}))\cdot \partial \mathbf{r}_2(s_{2i}, t_{2j})/\partial \mathbf{s}_{2i} = 0 \\ \text{and} \quad &(\mathbf{r}_{1ij} - \mathbf{r}_2(s_{2i}, t_{2j}))\cdot \partial \mathbf{r}_2(s_{2i}, t_{2j})/\partial \mathbf{t}_{2j} = 0 \end{aligned} \tag{2}$$

This system of non-linear equations was solved by the Newton–Raphson method to obtain $\mathbf{r}_{2ij}$ for each of the grid points $\mathbf{r}_{1ij}$. A new grid was set up around that point $\mathbf{r}_{1ij}$ for which the separation $\|\mathbf{r}_{1ij} - \mathbf{r}_{2ij}\|$ was the smallest. This new grid was finer than the original grid. This process was repeated several times with progressively smaller grids to locate the principal contact point.

(b) Calculating the principal curvatures and unit vectors

In order to determine the principal normal curvatures and directions of a surface $\mathbf{r}(s, t)$ at (s, t), define the matrices

$$[A] = \begin{bmatrix} \dfrac{\partial \mathbf{r}}{\partial s}\cdot\dfrac{\partial \mathbf{r}}{\partial s} & \dfrac{\partial \mathbf{r}}{\partial s}\cdot\dfrac{\partial \mathbf{r}}{\partial t} \\ \dfrac{\partial \mathbf{r}}{\partial t}\cdot\dfrac{\partial \mathbf{r}}{\partial s} & \dfrac{\partial \mathbf{r}}{\partial t}\cdot\dfrac{\partial \mathbf{r}}{\partial t} \end{bmatrix}$$

$$[B] = \begin{bmatrix} \mathbf{n}\cdot\dfrac{\partial^2 \mathbf{r}}{\partial s^2} & \mathbf{n}\cdot\dfrac{\partial^2 \mathbf{r}}{\partial s \partial t} \\ \mathbf{n}\cdot\dfrac{\partial^2 \mathbf{r}}{\partial t \partial s} & \mathbf{n}\cdot\dfrac{\partial^2 \mathbf{r}}{\partial t^2} \end{bmatrix} \tag{3}$$

where $\mathbf{n}$ is the unit normal vector,

$$\mathbf{n} = \frac{\dfrac{\partial \mathbf{r}}{\partial s} \times \dfrac{\partial \mathbf{r}}{\partial t}}{\left\| \dfrac{\partial \mathbf{r}}{\partial s} \times \dfrac{\partial \mathbf{r}}{\partial t} \right\|} \tag{4}$$

the matrix $[A]$ is the metric tensor for the surface and the matrix $[B]$ contains the normal curvature and twist values along the s and t directions.

If $\kappa^{(1)}$ and $\kappa^{(2)}$ are the two eigenvalues of the eigenproblem:

$$[B]\{\lambda\} = \kappa[A]\{\lambda\} \tag{5}$$

where

$$\{\lambda\} = \begin{Bmatrix} \lambda_1 \\ \lambda_2 \end{Bmatrix}$$

then $\kappa^{(1)}$ and $\kappa^{(2)}$ are the principal normal curvature values, and if the two corresponding eigenvectors $\{\lambda^{(1)}\}$ and $\{\lambda^{(2)}\}$ are normalized by

$$\{\lambda^{(1)}\}^{\mathrm{T}}[A]\{\lambda^{(1)}\} = 1$$

and

$$\{\lambda^{(2)}\}^{\mathrm{T}}[A]\{\lambda^{(2)}\} = 1 \tag{6}$$

then the unit vectors in the principal directions corresponding to the principal curvatures are

$$\mathbf{t}^{(1)} = \{\lambda^{(1)}\}^{\mathrm{T}} \begin{Bmatrix} \partial \mathbf{r}/\partial s \\ \partial \mathbf{r}/\partial t \end{Bmatrix}$$

and

$$\mathbf{t}^{(2)} = \{\lambda^{(2)}\}^{\mathrm{T}} \begin{Bmatrix} \partial \mathbf{r}/\partial s \\ \partial \mathbf{r}/\partial t \end{Bmatrix} \tag{7}$$

respectively. (The boldface symbols $\mathbf{t}_1^{(1)}$ and $\mathbf{t}_1^{(2)}$ should not be confused with the symbol t, which is one of the parameters defining the surface.)

(c) *Determining the relative principal curvature and directions*

Let $\kappa_1^{(1)}$, $\kappa_1^{(2)}$, $\mathbf{t}_1^{(1)}$, $\mathbf{t}_1^{(2)}$ be the two principal curvatures and two corresponding principal directions of surface no. 1 at a principal contact point $\mathbf{r}_1(s_1, t_1)$ and let $\kappa_2^{(1)}$, $\kappa_2^{(2)}$, $\mathbf{t}_2^{(1)}$, $\mathbf{t}_2^{(2)}$ be the two principal curvatures and two principal directions of surface no. 2 at $\mathbf{r}_2(s_2, t_2)$. Let θ be the angle between $\mathbf{t}_2^{(1)}$ and $\mathbf{t}_1^{(1)}$, defined as follows (Figure 2):

$$\theta = \tan^{-1}\left(\frac{\mathbf{t}_2^{(1)} \cdot \mathbf{t}_1^{(2)}}{\mathbf{t}_2^{(1)} \cdot \mathbf{t}_1^{(1)}}\right) \tag{8}$$

Let $K^{(1)}$ and $K^{(2)}$ be the two principal normal relative curvatures and let $\tau^{(1)}$ and $\tau^{(2)}$ be the unit vectors in the directions corresponding to these principal relative curvatures. Let Φ be the angle

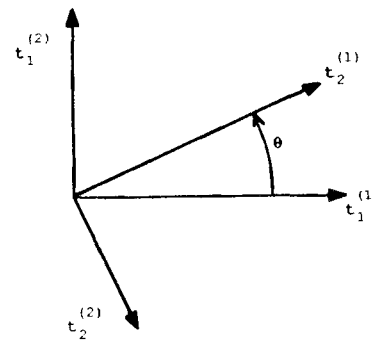


Figure 2. Directions of principal normal curvature

between $\tau^{(1)}$ and $t_1^{(1)}$. Then

$$\Phi = \frac{1}{2}\tan^{-1}\left[\frac{(\kappa_2^{(1)} - \kappa_2^{(2)})\sin(2\theta)}{(\kappa_1^{(1)} - \kappa_1^{(2)}) + (\kappa_2^{(1)} - \kappa_2^{(2)})\cos(2\theta)}\right] \tag{9}$$

$$\begin{aligned} K^{(1)} + K^{(2)} &= (\kappa_1^{(1)} + \kappa_1^{(2)}) + (\kappa_2^{(1)} + \kappa_2^{(2)}) \\ K^{(1)} - K^{(2)} &= (\kappa_1^{(1)} - \kappa_1^{(2)})\cos(2\Phi) + (\kappa_2^{(1)} - \kappa_2^{(2)})\cos(2(\Phi - \theta)) \end{aligned} \tag{10}$$

These equations yield the values of $K^{(1)}$ and $K^{(2)}$. The two directions for the principal normal relative curvatures are correspondingly

$$\begin{aligned} \tau^{(1)} &= t_1^{(1)}\cos\Phi + t_1^{(2)}\sin\Phi \\ \tau^{(2)} &= -t_1^{(1)}\sin\Phi + t_1^{(2)}\cos\Phi \end{aligned} \tag{11}$$

(d) Estimating the dimensions of the contact zone

After the relative curvatures have been obtained near the principal contact point, a rough estimate of the size of the contact zone can be obtained using Hertz's theory. Hertz's theory predicts an elliptic contact zone of a certain size. In general, the contact zone will not be an ellipse, and its size will be different from that predicted by Hertz's theory due to the fact that curvatures are not constant over the contact zone, and because the bodies are not elastic half spaces. Furthermore, the point of maximum contact pressure will not coincide with the principal contact point owing to the gross deformations of these bodies. Let F be the maximum possible normal force that is expected to fall on the contact zone around a principal contact point at $\mathbf{r}_1(s_1, t_1)$ and $\mathbf{r}_2(s_2, t_2)$. This value is obtained by assuming that all the load is borne by only one principal contact point and then using the condition of static equilibrium. The approximate dimensions of the contact ellipse for the load F are given by the lengths of the semi major and minor axes a and b, respectively:

$$\begin{aligned} a &= m\left(\frac{3\pi}{4}\frac{F(k_1 + k_2)}{(A + B)}\right)^{1/3} \\ b &= n\left(\frac{3\pi}{4}\frac{F(k_1 + k_2)}{(A + B)}\right)^{1/3} \end{aligned} \tag{12}$$

where

$$\begin{aligned} k_1 &= \frac{1 - \nu_1^2}{\pi E_1} \\ k_2 &= \frac{1 - \nu_2^2}{\pi E_2} \end{aligned} \tag{13}$$

are constants depending on material properties; A and B are determined by

$$\begin{aligned} (A + B) &= 0{\cdot}5|K^{(1)} + K^{(2)}| \\ (B - A) &= 0{\cdot}5|K^{(2)} - K^{(1)}| \end{aligned} \tag{14}$$

and the constants m and n are available in tabulated form[4] as functions of Ψ, where

$$\cos\Psi = (B - A)/(A + B) \tag{15}$$

(*e*) *Laying out the computational grids*

After lengths of the semi-axes of the contact ellipse have been determined, a grid of points is laid out on both the surfaces.

A grid of points $\mathbf{p}_{1ij}$ is set up near $\mathbf{r}_1(s_1, t_1)$ on surface no. 1 as follows:

$$\mathbf{p}_{1ij} = \mathbf{r}_1 + [(ia/n^{(1)})\boldsymbol{\tau}^{(1)} + (jb/n^{(2)})\boldsymbol{\tau}^{(2)}]\gamma \tag{16}$$

where

$$-n^{(1)} \leqslant i \leqslant n^{(1)}$$

and

$$-n^{(2)} \leqslant j \leqslant n^{(2)}$$

Here a and b are the estimated lengths of the ellipse semi-axes, $n^{(1)}$ and $n^{(2)}$ control the number of grid points and γ is a factor, usually about 1·4, required to make the grid larger than the contact ellipse size. The grid has to be made larger than the Hertzian contact ellipse to accomodate contact ellipses that are displaced from the principal contact point, and to allow for the fact that the actual contact zone might not be an ellipse of the predicted size. As long as the contact zone lies completely within this computational grid, the computed contact pressures are not very sensitive to the size of the grid. The grid must be made as small as possible, however, in order to achieve maximum resolution. The value of γ has to be modified if the contact zone is found to be truncated by the computational grid, or if the contact zone is found to be much smaller than the grid.

These points $\mathbf{p}_{1ij}$ will lie in a plane tangent to surface no. 1, as shown in Figure 1. Corresponding to each point $\mathbf{p}_{1ij}$, a grid point $\mathbf{r}_{1ij}$ is found on surface no. 1 which is closest to $\mathbf{p}_{1ij}$. This is done by minimizing $\|\mathbf{p}_{1ij} - \mathbf{r}_1(s_1, t_1)\|^2$. This is equivalent to solving the following pair of non-linear equations:

$$\begin{aligned}(\mathbf{p}_{1ij} - \mathbf{r}_1(s_1, t_1))\cdot \partial\mathbf{r}_1(s_1, t_1)/\partial s_1 = 0\\ (\mathbf{p}_{1ij} - \mathbf{r}_1(s_1, t_1))\cdot \partial\mathbf{r}_1(s_1, t_1)/\partial t_1 = 0\end{aligned} \tag{17}$$

for $(s_{1ij}, t_{1ij}) = (s_1, t_1)$. The solution yields $\mathbf{r}_{1ij} = \mathbf{r}_1(s_{1ij}, t_{1ij})$.

Similarly, a grid of points $\mathbf{p}_{2ij}$ is set up on the tangent plane to surface no. 2 at $\mathbf{r}_2(s_2, t_2)$:

$$\mathbf{p}_{2ij} = \mathbf{r}_2 + [(ia/n^{(1)})\boldsymbol{\tau}^{(1)} + (jb/n^{(2)})\boldsymbol{\tau}^{(2)}]\gamma \tag{18}$$

and the non-linear pair of equations:

$$\begin{aligned}(\mathbf{p}_{2ij} - \mathbf{r}_2(s_2, t_2))\cdot \partial\mathbf{r}_2(s_2, t_2)/\partial s_2 = 0\\ (\mathbf{p}_{2ij} - \mathbf{r}_2(s_2, t_2))\cdot \partial\mathbf{r}_2(s_2, t_2)/\partial t_2 = 0\end{aligned} \tag{19}$$

is solved for $(s_{2ij}, t_{2ij}) = (s_2, t_2)$. The solution yields $\mathbf{r}_{2ij} = \mathbf{r}_2(s_{2ij}, t_{2ij})$.

MATCHING SURFACE INTEGRAL AND FINITE ELEMENT SOLUTIONS

Let $\mathbf{u}(\mathbf{p}; \mathbf{q})$ denote the displacement vector at the location $\mathbf{q}$ due to a unit normal compressive force applied at the location $\mathbf{p}$ which is on the surface. The superscripts (si) and (fe) on a term will mean that the term has been calculated using surface integral formulae and a finite element model, respectively. Subscripts 1 and 2 will denote body no. 1 and body no. 2, respectively. When this subscript is omitted in an equation, the equation will be understood to apply to both the bodies.

Let $u(\mathbf{p};\mathbf{q}) = -\mathbf{u}(\mathbf{p};\mathbf{q})\cdot\mathbf{n}$ be the inward normal component of the displacement vector $\mathbf{u}(\mathbf{p};\mathbf{q})$, where $\mathbf{n}$ is the outward unit normal vector at the principal contact point.

The displacement $u(\mathbf{r}_{ij};\mathbf{r})$ of a field point $\mathbf{r}$ due to a load at the surface grid point $\mathbf{r}_{ij}$ can be expressed as

$$u(\mathbf{r}_{ij};\mathbf{r}) = (u(\mathbf{r}_{ij};\mathbf{r}) - u(\mathbf{r}_{ij};\mathbf{q})) + u(\mathbf{r}_{ij};\mathbf{q}) \tag{20}$$

where $\mathbf{q}$ is some location in the interior of the body, sufficiently removed from the surface (Figure 3). If the first two terms are evaluated using the surface integral formulae and the third term is computed from the finite element model, then we obtain the displacement estimate:

$$u(\mathbf{r}_{ij};\mathbf{r})(\mathbf{q}) = (u^{(\mathrm{si})}(\mathbf{r}_{ij};\mathbf{r}) - u^{(\mathrm{si})}(\mathbf{r}_{ij};\mathbf{q})) + u^{(\mathrm{fe})}(\mathbf{r}_{ij};\mathbf{q}) \tag{21}$$

The term in parentheses is the deflection of $\mathbf{r}$ with respect to the 'reference point' $\mathbf{q}$. This relative component is better estimated by a local deformation field based on the Bousinesq half space solution than by the finite element model. The gross deformation of the body due to the fact that it is not a half space will not significantly affect this term. On the contrary, the remaining term $u^{(\mathrm{fe})}(\mathbf{r}_{ij};\mathbf{q})$ is not significantly affected by local stresses at the surface. This is because $\mathbf{q}$ is chosen to be far enough beneath the surface. This term is therefore best computed using a finite element model of the body. The value $u(\mathbf{r}_{ij};\mathbf{r})(\mathbf{q})$ thus computed will, in general, depend on the location $\mathbf{q}$ because of the different values of the surface integral and finite element displacement fields there. The location is a so-called reference or 'matching' point. We would like to match the surface integral and finite element solutions not only at one point, but at a set of points belonging to a 'matching interface' (Figure 3). We will then be interested in that value for $u(\mathbf{r}_{ij};\mathbf{r})$ which will minimize the least squares deviation:

$$\int_{\Gamma} [u(\mathbf{r}_{ij};\mathbf{r}) - (u^{(\mathrm{si})}(\mathbf{r}_{ij};\mathbf{r}) - u^{(\mathrm{si})}(\mathbf{r}_{ij};\mathbf{q}) + u^{(\mathrm{fe})}(\mathbf{r}_{ij};\mathbf{q})]^2 \mathrm{d}\Gamma \tag{22}$$

where $\mathbf{q}$ varies over the reference surface Γ.

Another possibility, which lends itself better to spatial discretization, is to choose a value for $u(\mathbf{r}_{ij};\mathbf{r})$ which minimizes

$$\sum_{\mathbf{q}\in\Gamma} [u(\mathbf{r}_{ij};\mathbf{r}) - (u^{(\mathrm{si})}(\mathbf{r}_{ij};\mathbf{r}) - u^{(\mathrm{si})}(\mathbf{r}_{ij};\mathbf{q}) + u^{(\mathrm{fe})}(\mathbf{r}_{ij};\mathbf{q}))]^2 \tag{23}$$

where $\mathbf{q}$ varies over a grid of points $\mathbf{q}_{ij}$ laid out over the matching interface Γ (Figure 1). For programming convenience, points in this grid $\mathbf{q}_{ij}$ were chosen to lie half a finite element thickness below corresponding points in the surface grid $\mathbf{r}_{ij}$ (Figure 4). It remains to be determined how sensitive the final results are to this choice. Let N be the total number of points in the grid $\mathbf{q}_{ij}$.

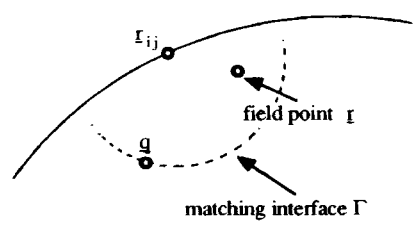


Figure 3. The matching interface

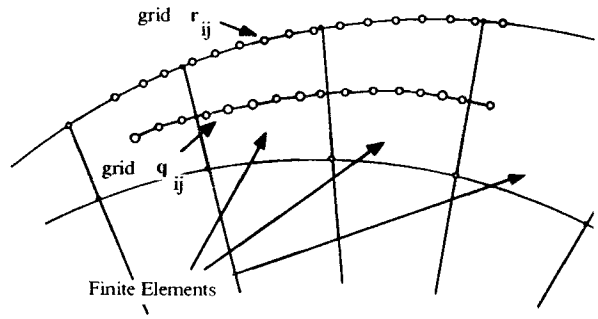


Figure 4. The sub-surface grid $\mathbf{q}_{ij}$

Then the value that minimizes the least square deviation above is

$$u(\mathbf{r}_{ij}; \mathbf{r}) = u^{(si)}(\mathbf{r}_{ij}; \mathbf{r}) + \frac{1}{N} \sum_{\mathbf{q}_{\alpha\beta}} [u^{(fe)}(\mathbf{r}_{ij}; \mathbf{q}_{\alpha\beta}) - u^{(si)}(r_{ij}; \mathbf{q}_{\alpha\beta})] \tag{24}$$

In order to obtain sufficient resolution of the contact stresses, the number of points in the grid $\mathbf{r}_{ij}$ will have to be very large, typically in the hundreds. Computation of all the terms of the type $u^{(fe)}(\mathbf{r}_{ij}; \mathbf{q}_{\alpha\beta})$ would involve hundreds of backsubstitutions. This would be prohibitively time consuming because of the complexity of the three-dimensional finite element model of the body. Furthermore, the finite element model cannot typically have an adequate degree of freedom at the surface to allow each of the terms $u^{(fe)}(\mathbf{r}_{ij}; \mathbf{q}_{\alpha\beta})$ to be independent of each other. Thus, evaluating each such term by a separate backsubstitution is probably also superfluous. A better method is to obtain $u^{(fe)}(\mathbf{r}_{\mathbb{i}(k)\mathbb{j}(k)}; \mathbf{q}_{\mathbb{i}(l)\mathbb{j}(l)})$ for a much smaller subset $\{\mathbf{r}_{\mathbb{i}(k)\mathbb{j}(k)}; k = 1, 2, \ldots, M\}$ of the grid $\{\mathbf{r}_{ij}\}$ as shown in Figure 5 and the corresponding subset $\{q_{\mathbb{i}(k)\mathbb{j}(k)}; k = 1, 2, \ldots, M\}$ of the grid $\{\mathbf{q}_{ij}\}$. If the number of points M in this restricted set of grid points is small, then all the terms $u^{(fe)}(\mathbf{r}_{\mathbb{i}(k)\mathbb{j}(k)}; \mathbf{q}_{\mathbb{i}(l)\mathbb{j}(l)})$ can be computed using only a small number M of backsubstitutions.

In the numerical examples to follow, M was 9, and these points were chosen from the grid as shown in Figure 5. The values of $u^{(fe)}(\mathbf{r}_{ij}; \mathbf{q}_{\alpha\beta})$ for the complete set of grid points can be obtained by using two-dimensional interpolants set up on the surface grid $\mathbf{r}_{ij}$ and the subsurface grid $\mathbf{q}_{ij}$, by the interpolation method:

$$u^{(fe)}(\mathbf{r}_{ij}; \mathbf{q}_{\alpha\beta}) = \sum_{k,l=1,M} u^{(fe)}(\mathbf{r}_{\mathbb{i}(k)\mathbb{j}(k)}; \mathbf{q}_{\mathbb{i}(l)\mathbb{j}(l)}) N_l(\alpha, \beta) N_k(i, j) \tag{25}$$

where the functions $N_k(i, j)$ are biquadratic functions of i and j:

$$N_k(i, j) = \sum_{\alpha,\beta=0,1,2} a_{k\alpha\beta} i^\alpha j^\beta \tag{26}$$

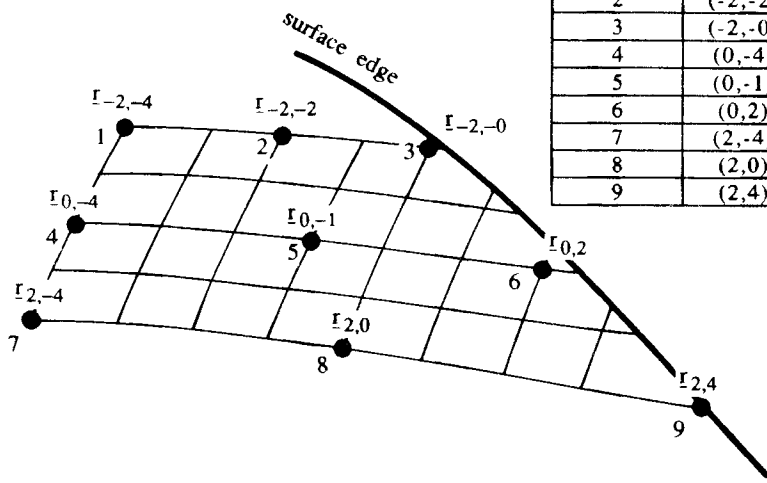
The coefficients $a_{k\alpha\beta}$ are chosen such that

$$N_k(\mathbb{i}(l), \mathbb{j}(l)) = \delta_{kl} \tag{27}$$

where δ_{kl} is the Kronecker delta.

The finite element formulation that was used to evaluate the terms $u^{(fe)}(\mathbf{r}_{\mathbb{i}(k)\mathbb{j}(k)}; \mathbf{q}_{\mathbb{i}(l)\mathbb{j}(l)})$ has been discussed in considerable detail in earlier papers,[10, 11] but a brief description has been included in the Appendix for convenience.

The computation of the surface integral terms $u^{(si)}(\mathbf{r}_{ij}; \mathbf{q}_{\alpha\beta})$ and $u^{(si)}(\mathbf{r}_{ij}; \mathbf{r})$ follows the methodology widely used in the tribology literature. If the point $\mathbf{r}$ lies on the surface of the body, then



k	(i(k),j(k))
1	(-2,-4)
2	(-2,-2)
3	(-2,-0)
4	(0,-4)
5	(0,-1)
6	(0,2)
7	(2,-4)
8	(2,0)
9	(2,4)

Figure 5. The computational grid $\mathbf{r}_{ij}$

define

$$l^{(1)} = |(\mathbf{r}_{ij} - \mathbf{r})\cdot\boldsymbol{\tau}^{(1)}|$$
$$l^{(2)} = |(\mathbf{r}_{ij} - \mathbf{r})\cdot\boldsymbol{\tau}^{(2)}| \tag{28}$$

to be the distance from $\mathbf{r}_{ij}$ to $\mathbf{r}$ along the principal axes $\boldsymbol{\tau}^{(1)}$ and $\boldsymbol{\tau}^{(2)}$, respectively, let

$$z = -(\mathbf{r}_{ij} - \mathbf{r})\cdot\mathbf{n} \tag{29}$$

and

$$\Delta x = \gamma(a/n^{(1)})$$
$$\Delta y = \gamma(b/n^{(2)}) \tag{30}$$

be the dimensions of each surface grid cell. The normal force F applied at $\mathbf{r}_{ij}$ is assumed to be evenly distributed over an area $\Delta x\Delta y$ centred around r_{ij}. The response to this distributed force can be obtained by integrating the Green's function (which is the Bousinesq solution). The result for $z = 0$, after integration is[3, 7]

$$u^{(si)}(\mathbf{r}_{ij};\mathbf{r}) = \frac{F}{2\pi G\Delta x\Delta y}\left[\left(y\sinh^{-1}\left(\frac{x}{|y|}\right) + x\sinh^{-1}\left(\frac{y}{|x|}\right)\right)^{x = l^{(1)} + \Delta x/2}_{x = l^{(1)} - \Delta x/2}\right]^{y = l^{(2)} + \Delta y/2}_{y = l^{(2)} - \Delta y/2} \tag{31}$$

If the point $\mathbf{r}$ is far enough below the surface, we can use the Bousinesq solution without integrating to obtain

$$u^{(si)}(\mathbf{r}_{ij};\mathbf{r}) = \frac{F}{2\pi Ed}[(1+\nu)z^2(z^2+d^2)^{-3/2} + 2(1-\nu^2)(z^2+d^2)^{-1/2}] \tag{32}$$

where $d = \|(\mathbf{r}_{ij} - \mathbf{r}) - \mathbf{zn}\|$.

The method described above is used to calculate all the terms $u_1(\mathbf{r}_{1ij};\mathbf{r}_{1kl})$ and $u_2(\mathbf{r}_{2ij};\mathbf{r}_{2kl})$ to build a compliance matrix. The contact force distribution over the grid and rigid body motions is

determined by setting up the contact equations using this compliance matrix and solving these contact equations by any of the numerous methods available in the literature. In the numerical examples described below, a method based on the Simplex algorithm of linear programming was used. Readers are referred to Vijayakar[10] for more details.

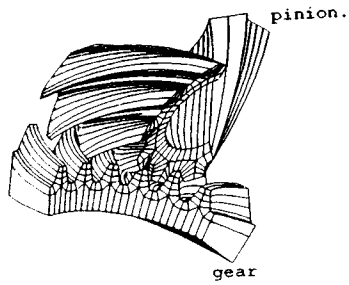


Figure 6. A six-tooth finite element model of the gear and the pinion

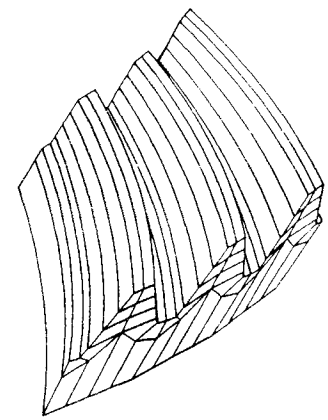


Figure 7. The three-tooth finite element model of the gear showing the active surfaces

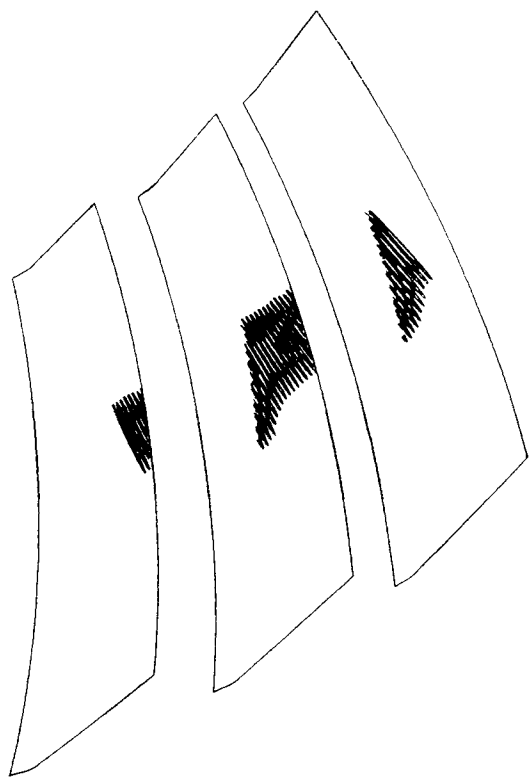


Figure 8. The locus of the contact zone as the gears roll against each other for a gear torque of 240 in-lb

NUMERICAL EXAMPLE

The case chosen as a numerical example is that of contacting hypoid gears. The surfaces of the teeth of these gears are created by a complicated cutting process. The cutting machines used have many kinematic settings. The settings are chosen such that the contact zone remains in the centre of the tooth surfaces as the gears roll against each other. A heuristic procedure is available to select the settings, but, in practice, these settings have to be selected after a tedious iterative process involving cutting and experimentally testing actual gears. Even so, it is very difficult to predict the actual contact stresses, fatigue life, kinematic errors and other design criteria, especially when not installed in ideal conditions. The contact stresses are so sensitive to the actual surface profile that conventional 3-D contact analysis is not feasible.

A sample 90° hypoid gear set from the rear axle of a commercial vehicle was selected. The gear ratio of this set was 41:11 and the axial offset was 1·5 in. The gear surfaces had been experimentally shown to be ideal for this particular gear ratio and axial offset. In other words, the contact zone was found to remain in the central portion of the gear teeth in the operational torque range. The object of this numerical study is to verify this by looking at the manner in which the contact pattern shifts when the gears are moved around from their ideal locations. Kinematic errors were also calculated.

The model was constructed by first generating values of co-ordinate normal vectors for points on the surface by simulating the cutting machines. The finite element description of the surface

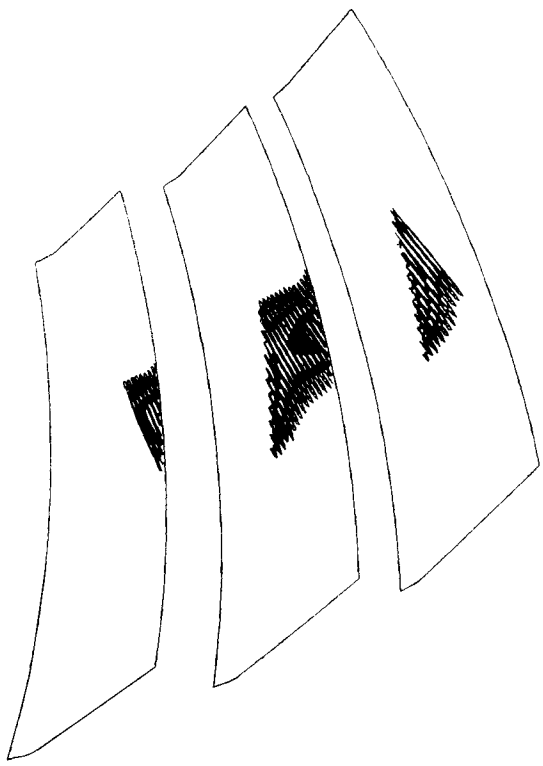


Figure 9. The locus of the contact zone as the gears roll against each other for a gear torque of 480 in-lb

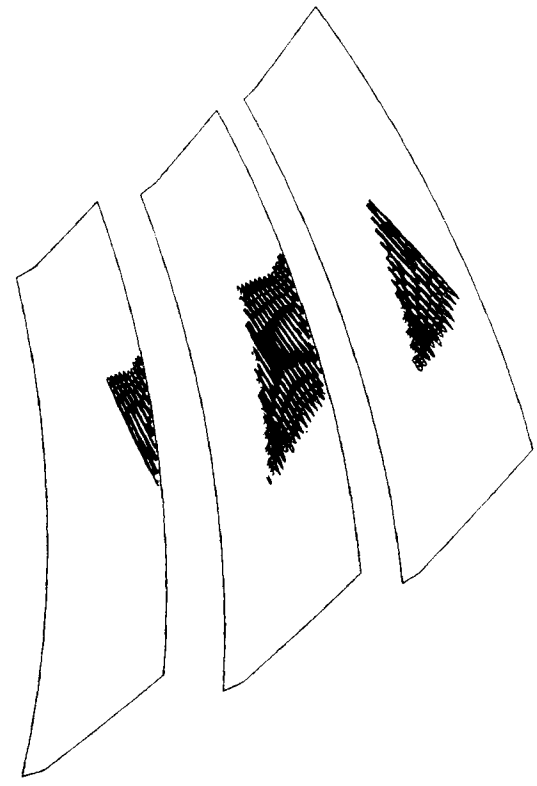


Figure 10. The locus of the contact zone as the gears roll against each other for a gear torque of 960 in-lb

was then created by fitting tenth order truncated Chebyshev series approximations to these data. The interior portions of the finite element were created semi-automatically. Only a sector containing three teeth of each gear was modelled, with each tooth being identical. The gear (gear no. 1) and the pinion (gear no. 2, the smaller gear) were then oriented in space as per the assembly drawings, and the analysis was carried out for each individual time step. Figure 6 shows a six-tooth gear and pinion model. Sectoral symmetry is used to generate stiffness matrices from the stiffness matrix of one tooth. For this particular gear set, a three-tooth model suffices because at the most two teeth contact at a time. Figure 7 shows the surfaces of the three-tooth gear. Figures 8, 9 and 10 show the contact pattern (which is the locus of the contact zone as the gears roll against each other) for a gear torque of 240, 480 and 960 in-lb, respectively. Figures 11 and 13 show views of the contact zone with contact pressure contours on the gear for two particular angular positions. Figures 12 and 14 show magnified views of the contact zone for these two positions. They show contours of normal contact pressures on the surfaces. Computational grids of 11 × 25 cells were used on these surfaces to obtain the pressure distributions. Another useful piece of information that can be obtained is the kinematic transmission error, which is the deviation of the motion of the gears from their ideal angular motion. Figure 15 shows the transmission error in radians for three different torque levels (0, 480 and 960 in-lb). These

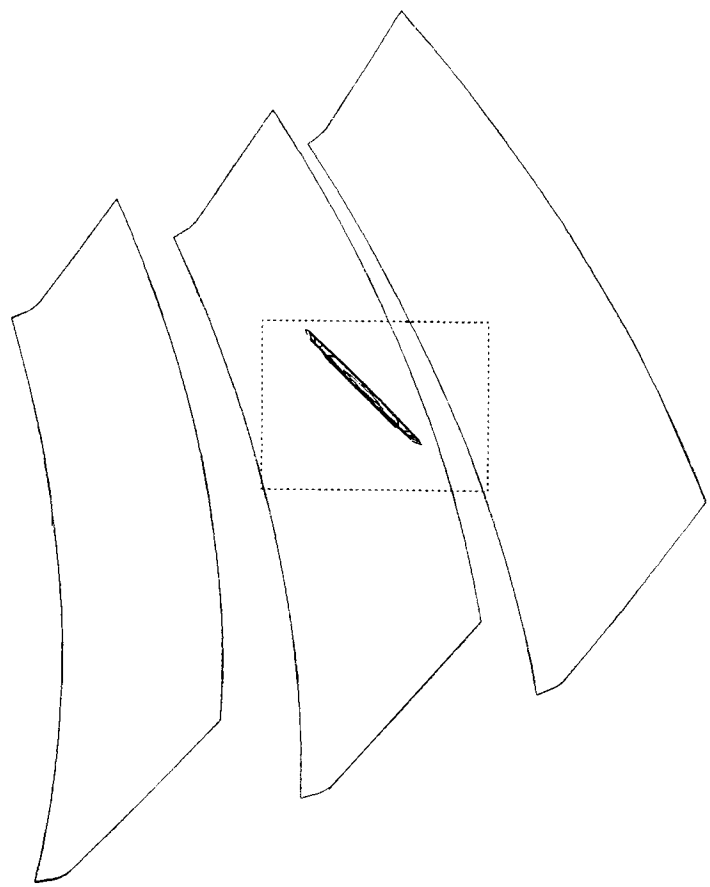


Figure 11. Contact stress contours for position 1

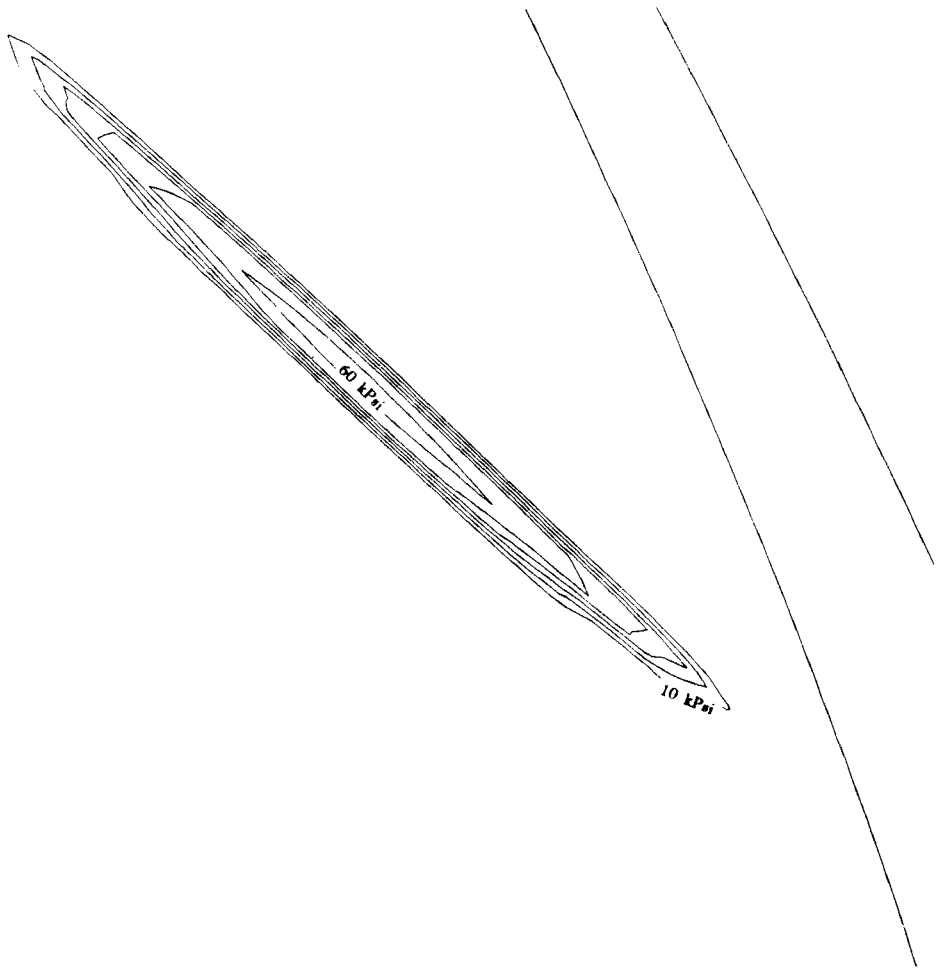


Figure 12. Contact stress contours for position 1

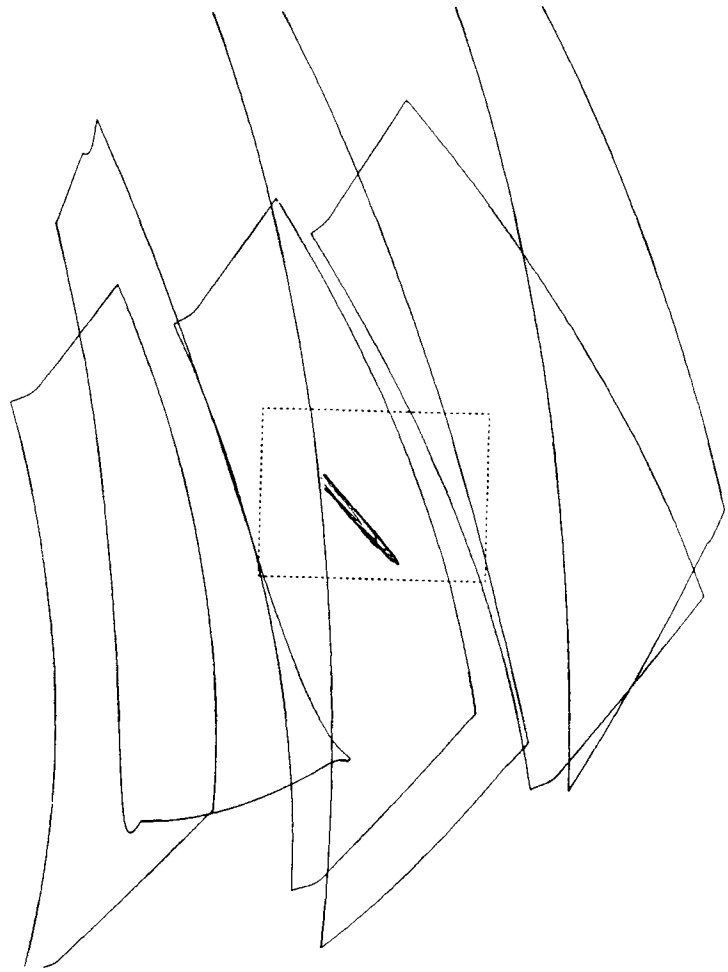


Figure 13. Contact stress contours for position 2

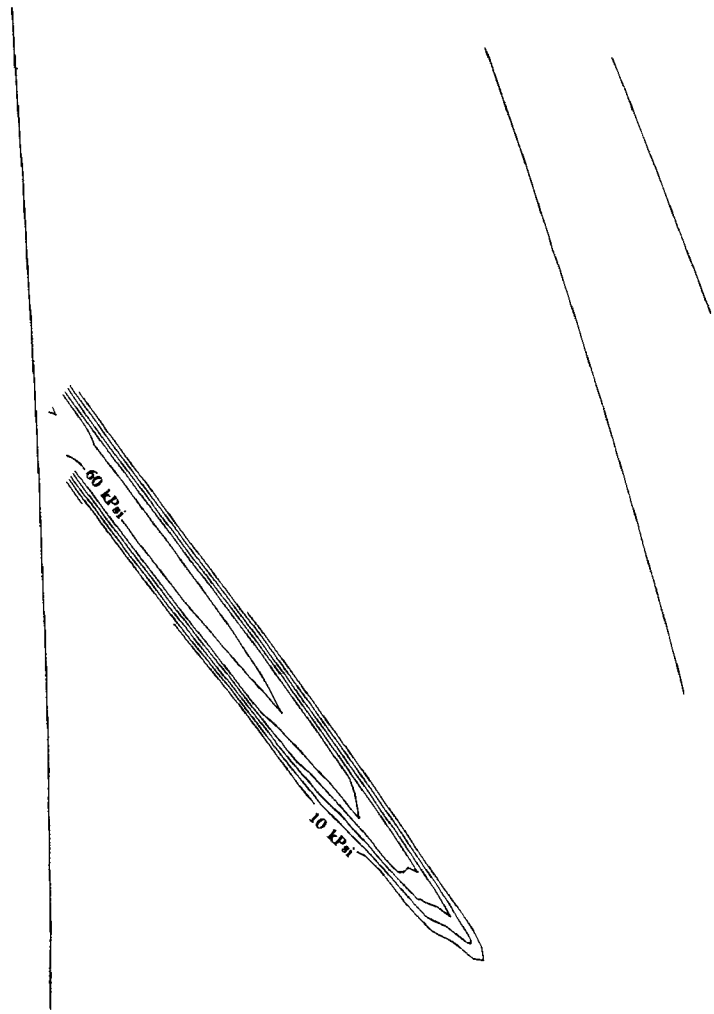


Figure 14. Contact stress contours for position 2

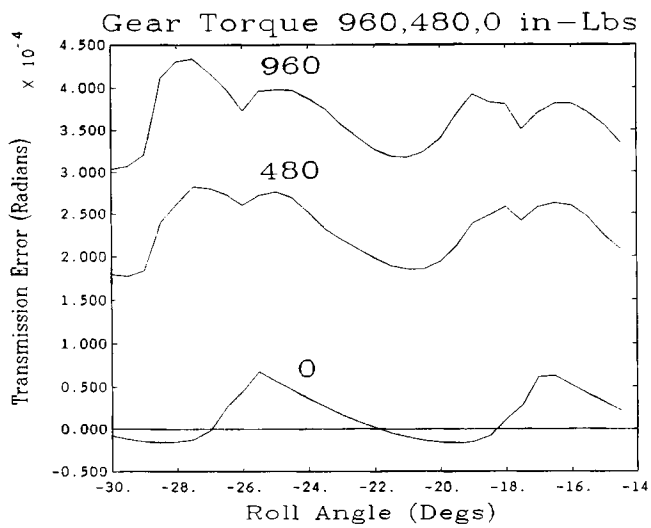


Figure 15. The transmission error curves obtained for 960, 480 and 0 in-lb of torque applied to the gear

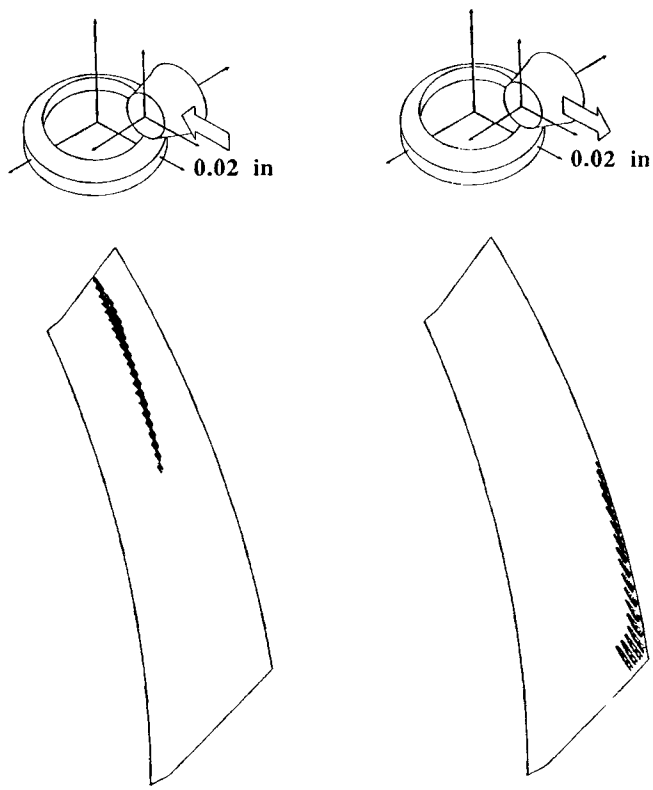


Figure 16. The effect of an X translation on the contact pattern

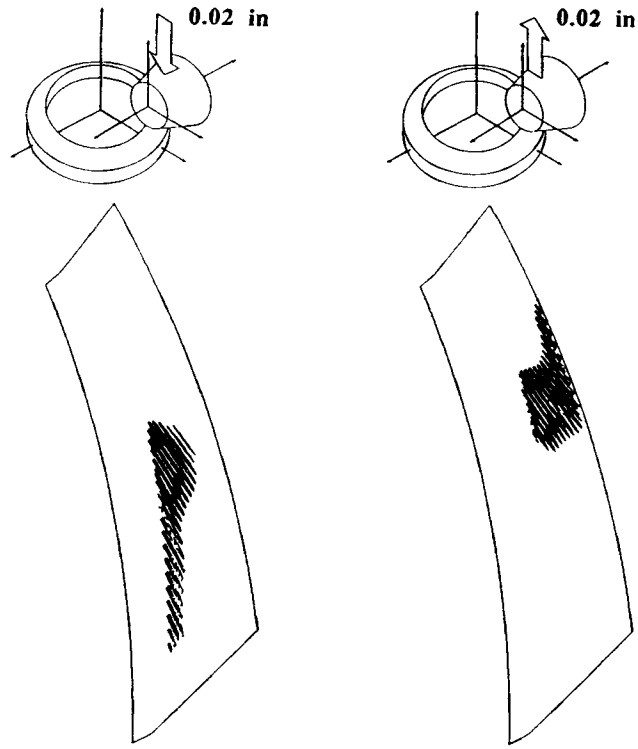


Figure 17. The effect of an Y translation on the contact pattern

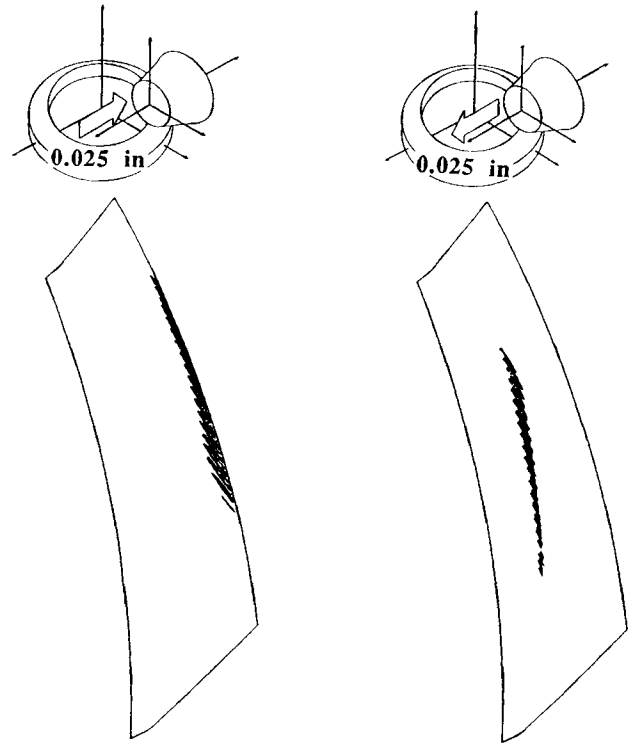


Figure 18. The effect of a Z translation on the contact pattern

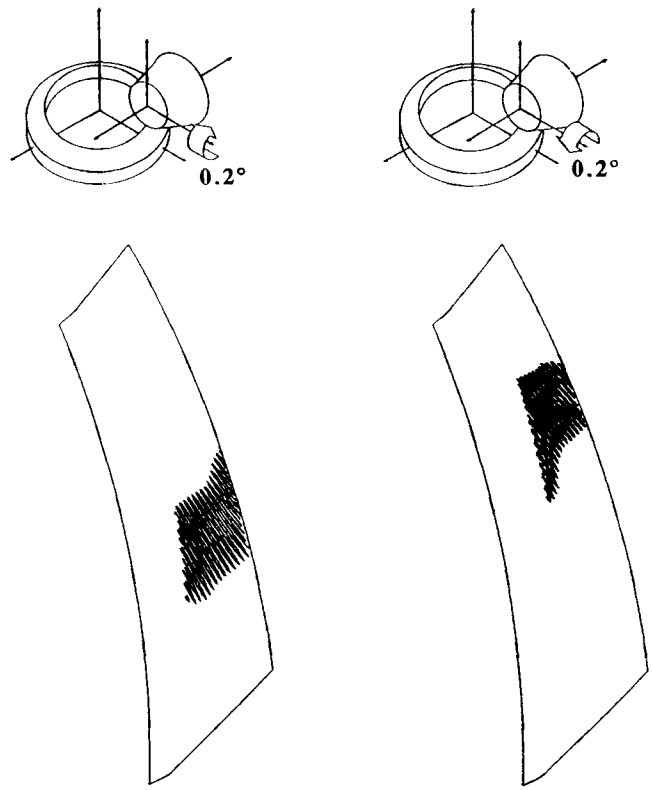


Figure 19. The effect of an X rotation on the contact pattern

transmission error curves are remarkably similar to those measured experimentally by other researchers.[5] Finally, the position of the pinion was perturbed slightly from the design location, and Figures 16 to 19 show the contact patterns that were obtained. When compared to the contact pattern for the unperturbed position in Figure 7, it shows that the best contact pattern does indeed occur at the designed position, lending credence to the notion that an analysis of the kind described in this paper has the potential to be used in the design process itself.

CONCLUSIONS

Using a combination of finite element and surface integral methods seems to be, in the author's opinion, the most practical method of modelling stiffness behaviour of contacting bodies. Used along with an efficient algorithm for solving contact equations, one can predict contact stress distributions and deformations in more realistic detail than otherwise possible.

Experimental data are not available at the present time for comparison and verification, owing to the difficulties of measuring contact pressures over very small areas. Future efforts will be directed towards experimentally estimating the accuracy of the method, and the development of a rational strategy to determine parameters like the optimal depth of the matching interface. An effort will be made to incorporate some of the more elaborate elastohydrodynamic contact models that exist.[6]

ACKNOWLEDGEMENTS

The author would like to thank the sponsors of the Gear Dynamics and Gear Noise Research Laboratory of The Ohio State University and the Gleason Memorial Fund for their financial support, and also the Advanced Design Methods Laboratory of The Ohio State University for the use of its computational facilities.

APPENDIX

Finite element representation of the contacting bodies

The displacements $\mathbf{u}^e(\xi, \eta, \zeta)$ and co-ordinates $\mathbf{x}^e(\xi, \eta, \zeta)$ are approximated within an element by

$$\mathbf{u}^e(\xi, \eta, \zeta) = \sum_{i=1}^{\tilde{m}^e} \sum_{j=0}^{\tilde{n}_i^e} \mathbf{u}_{ij}^e \tilde{N}_i(\xi, \eta) \tilde{T}_{ij}(\zeta)$$

$$\mathbf{x}^e(\xi, \eta, \zeta) = \sum_{i=1}^{\bar{m}^e} \sum_{j=0}^{\bar{n}_i^e} \mathbf{x}_{ij}^e \bar{N}_i(\xi, \eta) \bar{T}_{ij}(\zeta)$$

where e is the element index, $\tilde{m}^e$ and $\bar{m}^e$ are the numbers of conventional two-dimensional shape functions $\tilde{N}_i(\xi, \eta)$ and $\bar{N}_i(\xi, \eta)$ within the two-dimensional domain $\xi, \eta \in [-1, +1]$ of the element used to represent the displacements and co-ordinates, respectively. $\tilde{T}_{ij}(\zeta)$ and $\bar{T}_{ij}(\zeta)$ are the first few members of a complete set of functions defined in $\zeta \in [-1, +1]$.

The two-dimensional shape functions $\tilde{N}_i(\xi, \eta)$ used for displacements are conventional shape functions. In this study, they were quadratic isoparametric shape functions corresponding to an element with between four and nine nodes.

For co-ordinate interpolation, two different kinds of elements were used, depending on whether or not part of the element's surface was along the active contacting surface of the body. For interior elements, the shape functions $\bar{N}_i(\xi, \eta)$ used were those of a four noded linear two-dimensional finite element. For elements that share part of the contacting surface, a special element (Figure 20) formulation was used that allows for a C_1 continuous representation of the surface and accuracies within the prescribed surface tolerances.

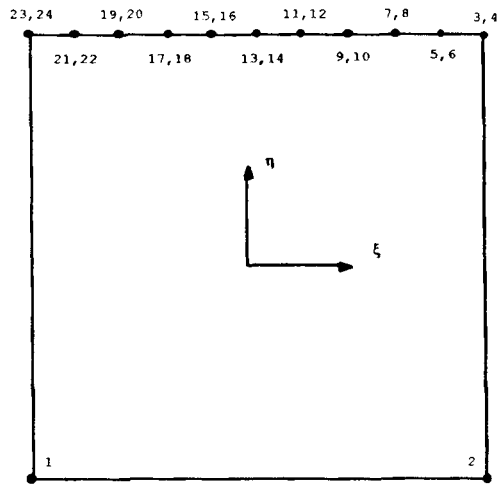


Figure 20. The two-dimensional element used for co-ordinate representation at the surface

Define

$$f_1(s) = (1 - s)/2$$
$$f_2(s) = (1 + s)/2$$

and let

$$H_1(s) = (s^3 - 3s + 2)/4$$
$$H_2(s) = (-s^3 + 3s + 2)/4$$
$$H_3(s) = (s^3 - s^2 - s + 1)/4$$
$$H_4(s) = (s^3 + s^2 - s + 1)/4$$

Here $H_1(s)$, $H_2(s)$, $H_3(s)$ and $H_4(s)$ are the Hermite cubic shape functions used in beam elements that allow C_1 continuous one-dimensional interpolation. Figure 1 shows the two-dimensional element that uses the following 24 shape functions defined for $\xi, \eta \in [-1, +1]$:

$$\bar{N}_1(\xi, \eta) = f_1(\xi) f_1(\eta)$$
$$\bar{N}_2(\xi, \eta) = f_2(\xi) f_1(\eta)$$
$$\bar{N}_3(\xi, \eta) = \begin{cases} H_2(s_{10}(\xi)) f_2(\eta) & 0{\cdot}8 \leqslant \xi \\ 0 & \xi < 0{\cdot}8 \end{cases}$$
$$\bar{N}_4(\xi, \eta) = \begin{cases} H_4(s_{10}(\xi)) f_2(\eta) & 0{\cdot}8 \leqslant \xi \\ 0 & \xi < 0{\cdot}8 \end{cases}$$

where

$$s_{10}(\xi) = 2(\xi - 0{\cdot}8)/0{\cdot}2 - 1{\cdot}0$$
$$\bar{N}_5(\xi, \eta) = \begin{cases} H_1(s_{10}(\xi)) f_2(\eta) & 0{\cdot}8 \leqslant \xi \\ H_2(s_9(\xi)) f_2(\eta) & 0{\cdot}6 < \xi \leqslant 0{\cdot}8 \\ 0 & \xi < 0{\cdot}6 \end{cases}$$
$$\bar{N}_6(\xi, \eta) = \begin{cases} H_3(s_{10}(\xi)) f_2(\eta) & 0{\cdot}8 \leqslant \xi \\ H_4(s_9(\xi)) f_2(\eta) & 0{\cdot}6 < \xi \leqslant 0{\cdot}8 \\ 0 & \xi < 0{\cdot}6 \end{cases}$$

where

$$s_9(\xi) = 2(\xi - 0{\cdot}6)/0{\cdot}2 - 1{\cdot}0$$
$$\bar{N}_7(\xi, \eta) = \begin{cases} H_1(s_9(\xi)) f_2(\eta) & 0{\cdot}6 \leqslant \xi < 0{\cdot}8 \\ H_2(s_8(\xi)) f_2(\eta) & 0{\cdot}4 < \xi \leqslant 0{\cdot}6 \\ 0 & \text{otherwise} \end{cases}$$
$$\bar{N}_8(\xi, \eta) = \begin{cases} H_3(s_9(\xi)) f_2(\eta) & 0{\cdot}6 \leqslant \xi < 0{\cdot}8 \\ H_4(s_8(\xi)) f_2(\eta) & 0{\cdot}4 < \xi \leqslant 0{\cdot}6 \\ 0 & \text{otherwise} \end{cases}$$

where

$$s_8(\xi) = 2(\xi - 0{\cdot}4)/0{\cdot}2 - 1{\cdot}0$$

and so on, until

$$\bar{N}_{23}(\xi, \eta) = \begin{cases} H_1(s_1(\xi))f_2(\eta) & -1{\cdot}0 < \xi \leqslant -0{\cdot}8 \\ 0 & \text{otherwise} \end{cases}$$

$$\bar{N}_{24}(\xi, \eta) = \begin{cases} H_3(s_1(\xi))f_2(\eta) & -1{\cdot}0 < \xi \leqslant -0{\cdot}8 \\ 0 & \text{otherwise} \end{cases}$$

where

$$s_1(\xi) = 2(\xi + 1{\cdot}0)/0{\cdot}2 - 1{\cdot}0$$

The functions $\bar{T}_{ij}(\zeta)$ and $\tilde{T}_{ij}(\zeta)$ were based on the Chebyshev polynomials:

$$\bar{T}_{in}(\zeta) = \tilde{T}_{in}(\zeta) = \begin{cases} (1 - \zeta)/2 & \text{for } j = 0 \\ (1 + \zeta)/2 & \text{for } j = 0 \\ \tau_n(\zeta) - 1 & \text{if } n > 1 \text{ and } n \text{ is odd} \\ \tau_n(\zeta) - \zeta & \text{if } n > 1 \text{ and } n \text{ is even} \end{cases}$$

where $\tau_n(\zeta)$ is the Chebyshev polynomial of order n defined on the interval $[-1, +1]$.

REFERENCES

1. K. J. Bathe and A. Chowdhury, 'A solution method for planar and axisymmetric contact problems', *Int. j. numer. methods eng.*, **21**, 65–88 (1985).
2. A. Chowdhury and K. J. Bathe, 'A solution method for static and dynamic analysis of three-dimensional contact problems with Friction', *Comp. Struct.*, **24**, 855–873 (1986).
3. J. M. de Mul, J. J. Kalker and B. Fredriksson, 'The contact between arbitrarily curved bodies of finite dimensions', *J. Tribology ASME*, **108**, 140–148 (1986).
4. Murray Kornhauser, 'A note on elastic surface deformation', *J. Appl. Mech. ASME*, **18**, 251–252 (1951).
5. T. Krenzer, 'Gear sound analysis—I', Sheet No. 9, *Engineering Report No. 4731*, Gleason Works, Rochester, N.Y., 1973.
6. W. T. Lai and H. S. Cheng, 'Computer simulation of elastic rough contacts', *ASLE Trans.*, **28**, 172–180 (1985).
7. A. A. Lubrecht, W. E. ten Napel and R. Bosma, 'Multigrid, an alternative method of solution for two-dimensional elastohydrodynamically lubricated point contact calculations', *J. Tribology ASME*, **109**, 437–443 (1987).
8. J. Schwartz and E. Harper, 'On the relative approach of two-dimensional bodies in contact', *Int. J. Solids Struct.*, **7**, 1613–1626 (1971).
9. J. Seabra and D. Berthe, 'Influence of surface waviness on the normal pressure distribution in the Hertzian contact', *J. Tribology ASME*, **109**, 462–470 (1987).
10. S. Vijayakar, H. Busby and D. Houser, 'Linearization of multibody frictional contact problems', *Comp. Struct.*, **29**, 569–576 (1988).
11. S. Vijayakar, H. Busby and L. Wilcox, 'Finite element analysis of three-dimensional conformal contact with friction', *Comp. Struct.*, **33**, 49–62 (1989).
12. C. Weber, 'The deformations of gears and their load carrying capacity', *Research Report 3*, British Department of Scientific and Industrial Research, 1949.
13. C. Weber, 'Formanderung und Profilrucknahme bei Gerad- und Schragverzahnten Radern', *Schriftenreihe Antriebstechnik*, Heft II (1953).