Involute Inspection Methods and Interpretation of Inspection Results

Yefim Kotlyar

Introduction

What is so unique about gear manufacturing and inspection? Machining is mostly associated with making either flat or cylindrical shapes. These shapes can be created by a machine's simple linear or circular movements, but an involute curve is neither a straight line nor a circle. In fact, each point of the involute curve has a different radius and center of curvature. Is it necessary to go beyond simple circular and linear machine movements in order to create an involute curve? One of the unique features of the involute is the fact that it can be generated by linking circular and linear movements. This uniqueness has become fertile soil for many inventions that have simplified gear manufacturing and inspection. As is the case with gear generating machines, the traditional involute inspection machines take advantage of some of the involute properties. Even today, when computers can synchronize axes for creating any curve, taking advantage of involute properties can be very helpful. It
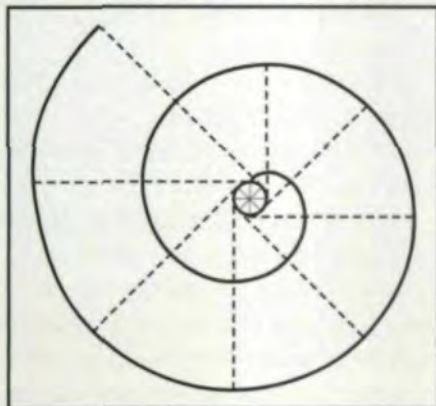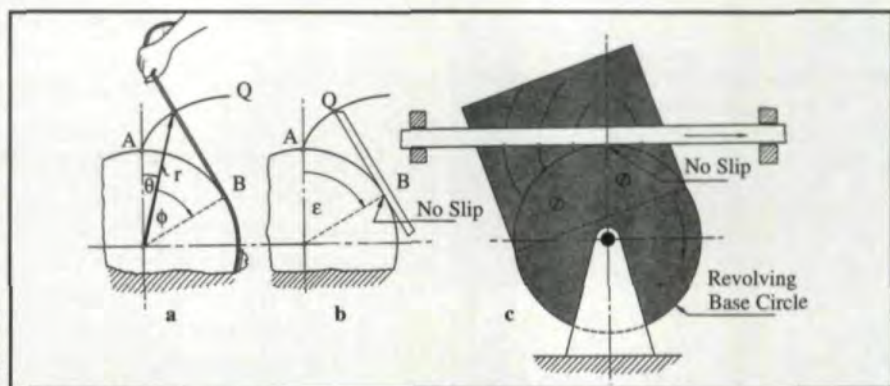

Fig. 2 — Methods for generating an involute curve: a) the string method; b) a beam rolling around a fixed base circle; c) beam and base circle rolling with each other without slip.

can simplify synchronization of machine movements and reduce the number of variables to monitor.

Involute Definition, Geometric Properties and Involute Function

The involute curve is a spiral beginning at the base circle and having an infinite number of equidistant coils (Fig. 1). Only a small portion of the innermost coil has been utilized in practical applications. The easiest way to visualize this is by describing the way it can be generated. The involute curve can be generated by a point on a tightly held, inextensible and extremely thin thread that is unwound from a fixed circle, called the base circle (Fig. 2a). This method is called the string method. An involute can also be generated by a beam rolling around a fixed base circle (Fig. 2b) or by a beam and base circle rolling with each other without slip (Fig. 2c). All these principles are used in gear generating and inspection machines.

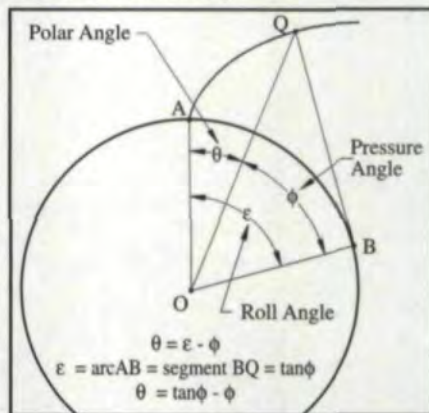Important geometric properties of the involute curve can be derived from its

$$\theta = \varepsilon - \phi$$
$$\varepsilon = \text{arcAB} = \text{segment BQ} = \tan\phi$$
$$\theta = \tan\phi - \phi$$

Fig. 3 — Geometric properties of the involute.

generation. Some of these properties are used either for the inspection machine movements or referenced in the inspection results.

• The line tangent to the base circle, drawn from any point of the involute, is always perpendicular to the involute curve (see Fig. 3).

• The segment of the tangent line, QB, is the radius of curvature of the involute for the point Q. Points where the string separate from the base circle are instantaneous centers of the involute curvature.
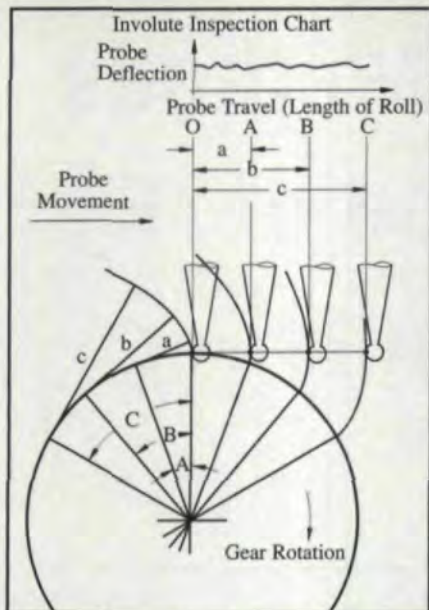

Fig. 1 — The basic involute curve.

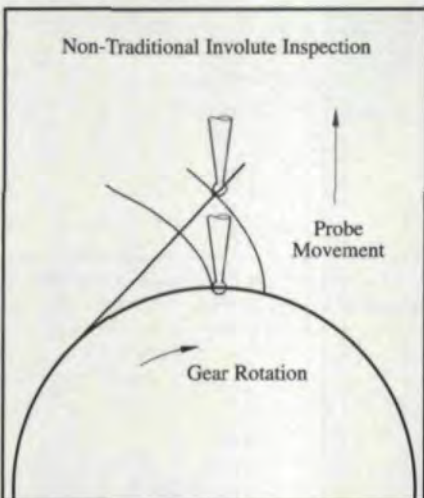Fig. 4 — Basic mechanics of an involute inspection process using a probe.

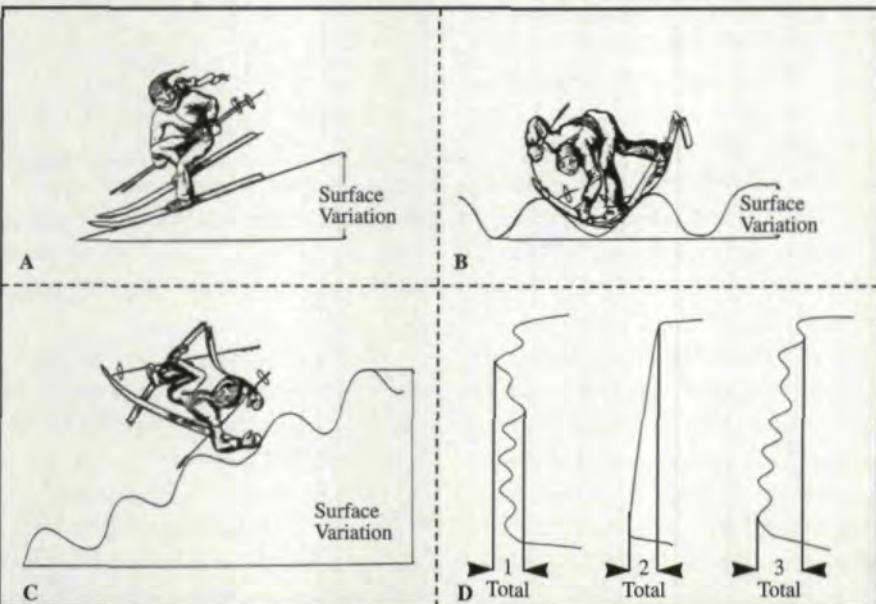

Fig. 5 — Mechanics of non-traditonal involute inspection.

Fig. 6 — Surface variation from the horizontal plane.

• For any point on the involute, arc length *AB*, contained by the beginning of the involute and the point of tangency, is equal to the length of the line segment *Q*B tangent to the base circle.

Involute Function

Let the involute or polar angle be θ, the pressure angle be ϕ and the roll angle be ε. Then let us assume for simplicity that the base radius equals 1 unit of linear measurement. In this case, the length of an arc equals the angular measurement using radians as measuring units. Therefore, $\theta = \varepsilon - \phi$, where $\varepsilon = \text{arc}AB = \text{segment } BQ = OB \cdot \tan\phi = 1 \cdot \tan\phi = \tan\phi$. The involute function can be derived by replacing ε with its function of ϕ: $\theta = \tan\phi - \phi$.

Most analytical gear inspection machines use these involute geometry properties:

• Length of roll *QB* equals the length of arc of roll *AB*.

• The line tangent to the base circle is always perpendicular to the involute curve.

Principles of Traditional Mechanical or CNC Involute Inspection and Resulting Charts

The Mechanics of Machine Movement. Traditional involute inspection concepts are based on combining the linear motion of the probe carrier and the rotational motion of the gear. This combined movement generates an involute path for the probe relative to the gear profile. While the probe moves along the path that is tangent to the base circle, the

distance equals the length of roll *a*, and the gear rotates the angle *A* (Fig.4).

One important beneficial distinction of the traditional involute inspection method is the unchanging probe contact point throughout the entire probe travel (Fig. 4). This unchanging contact point simplifies the inspection process by reducing the number of variables that need to be monitored.

The probe deflection represents a deviation of the gear profile from the involute curve. If the gear profile is a perfect involute, the probe deflection would stay constant throughout the entire movement, and the resulting inspection chart would be a straight horizontal line. A deviation from this straight line would constitute the profile error.

Involute Inspection Charts. An involute inspection chart is scaled proportionately to the length or angle of roll. The *X* coordinate (probe travel) represents length or angle of roll. The *Y* coordinate represents profile deviation from the perfect involute in the direction normal to the involute curvature.

It is important to reiterate that the traditional involute inspection chart is not proportional to the diameter, nor is it proportional to the length of involute curve.

Non-Traditional Involute Inspection

With the proliferation of coordinate measuring machines, other involute inspection methods have come into being. Some CMMs use the traditional method, but some don't. Nevertheless, the inspection results are presented in the old fashioned way—profile tracing is scaled proportionately to the length of angle of roll, as shown in the upper section of Fig. 4.

Machines that do not use the traditional method include

• CMMs without rotary tables. The probe contours a fixed gear.

• CMMs with rotary tables, but without tangential slides.

The principle difference between non-traditional and traditional machines is the fact that non-traditional machines have three axes instead of four. A fewer number of axes makes one part of the machine less expensive; however, it also creates an additional burden in

another area of the machine. In both non-traditional cases, in addition to two moving machine axes, the system has to keep track of one extra variable—the contact point of the probe. Thus, the machine cannot take full advantage of involute properties for reducing the number of variables to monitor during involute inspection.

Fig. 5 depicts the involute inspection principle for the machine without a tangential slide. X & Y coordinates of the probe contact are continuously changing as the probe moves from root to tip (See Fig. 5). To make matters worse, in the case of helical gears, X, Y and Z coordinates of the probe contact are continuously changing.

The advantages of non-traditional machines are that they have fewer axes, and their 3-dimensional probes give them the potential for adding non-gearing inspection capabilities to the machines.

The disadvantages of using these non-traditional machines include the need to monitor the extra variables during involute inspection, which can make the systems either less accurate or more expensive to develop, and the requirement for 3-dimensional probes, which add a significant cost to the apparatus.

Some Common Principles of Surface Evaluation

How do people analyze, qualify and quantify the surface deviation from desired conditions? What do we mean by "profile error"? Is this the amount of error or the shape of the error or both? There are situations in which one number for defining involute error is not sufficient to quantify and qualify the error.

Let's introduce three definitions: slope error, form error and total error. The drawings in Fig. 6 help to illustrate the differences between these three concepts. Fig. 6a shows surface variation from the horizontal plane. Is it a lot or a little? For a farmer it may be lot, but for a skier, it may not be enough. Fig. 6b shows a different type of surface variation. Even for some skiers, it may be too much. In reality however, people frequently deal with a combination of the kinds of surface errors shown in a and b. The situation is more like the one shown in Fig. 6c. The errors may have the same value, but a different appearance, as shown in the Fig. 6d.

Determination of Total, Slope and Form Errors

The total surface deviation from an ideal condition can be broken down into slope and form errors (Fig. 7). There are

various techniques for isolating these errors. The most popular, and probably the most accurate, is the "least squares" method. This method is based on determining the "best fit" line that segments the curve into two approximately equal areas on either side of that line (see Fig. 7). Deviation of the best fit line from the ideal position (the vertical in Fig. 7), is called the slope error. Deviation of the inspection curve from the best fit line is called the form error.
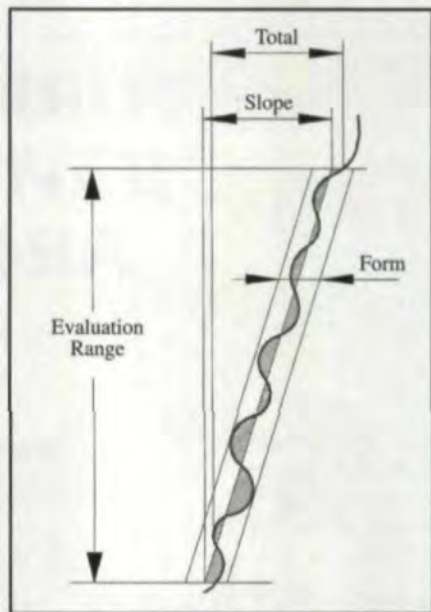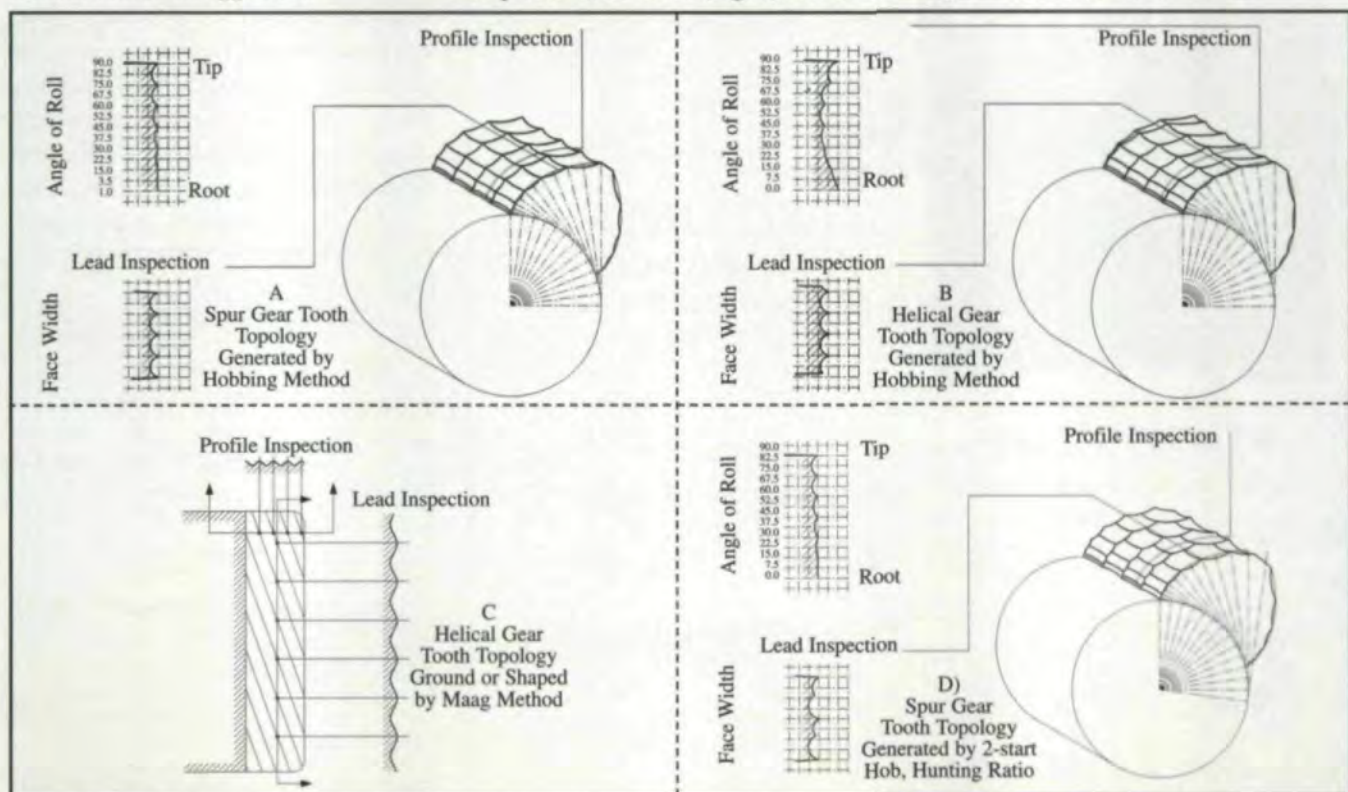

Fig. 7 — Total, slope and form errors.

Fig. 8 — Spur gear tooth topology.

The breakdown of total error into form and slope components is applicable to both involute and lead inspection. Because slope and form errors come from different sources, isolating and assigning a value to each error component is very helpful for finding the largest contributor to the tooth surface inaccuracy.

Slope error sources include

• Lead. The wrong machine settings for the helix angle.

• Involute. Wrong hob pressure angle or wrong rake angle induced during hob sharpening.

Form error sources include

• Lead. Excessive feed rate.

• Involute. Excessive hob runout or hob gash index error or excessive hob lead error or insufficient number of hob gashes.

The breakdown of the total error into the slope and form error components can be invaluable for determining exact machine or tool adjustments, thus eliminating time-consuming trial and error techniques. For example, lead or taper adjustment on a CNC hobbing machine can be determined accurately with the average lead slope errors.

Tooth Surface Patterns Created by Various Manufacturing Processes

Every gear manufacturing method creates a certain gear topology that, during the inspection, translates mostly into form errors. Some examples of tooth topology are shown in Figs. 8a–d. Note that all these examples display involute and lead errors despite the fact that these gears were manufactured under ideal conditions of machine, tool, fixture and blank. These errors are referred to as inherent errors introduced by the process principle.

Even the most accurate hobs and machines can create greater than allowed lead and involute errors. Understanding the tooth topology helps to differentiate between the inherent errors introduced by the process principle and errors induced by the process variables (machine, cutting tool, fixture and blank inaccuracies). Frequently, determining the error source is a bigger challenge than the elimination of that source.

Common Profile Modification

Commonly gears are designed with tip and/or root relief (Fig. 9). The numerical evaluation of such profiles becomes more complicated. There are various computerized techniques available for evaluation of profile modifications. Some people use a comparison with an ideal curve. Some evaluate various portions of the tracing separately; some evaluate crown; some evaluate hollow;
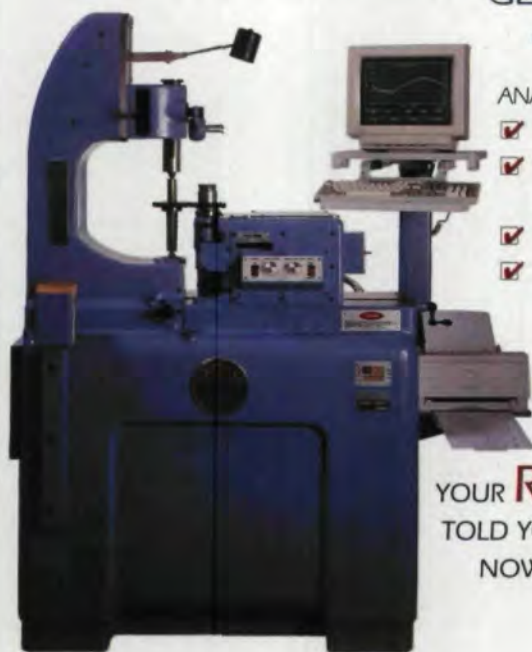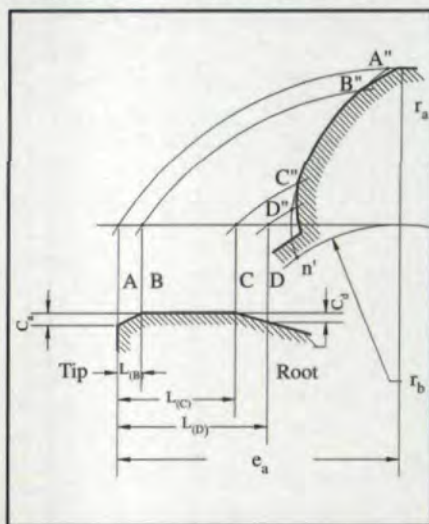
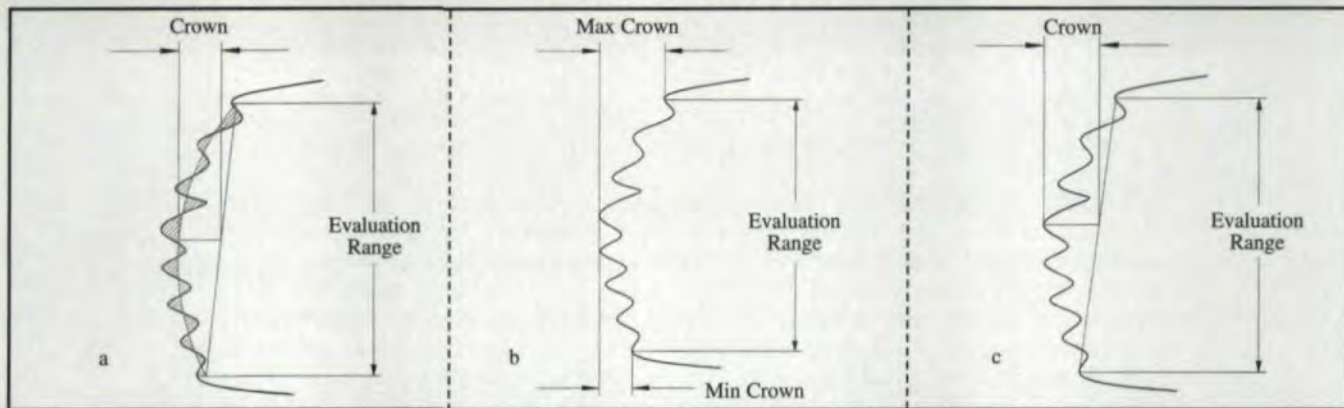Fig. 9 — Involute geometry of a gear designed with tip and root relief.

Fig. 10 — Computized techniques for crown evaluation: a) best fit curve method; b) comparison of both ends of tracing with high point; c) comparison of a line connecting tracing ends with high point.
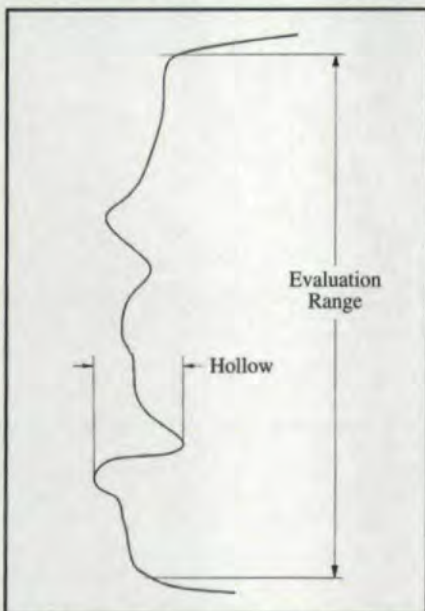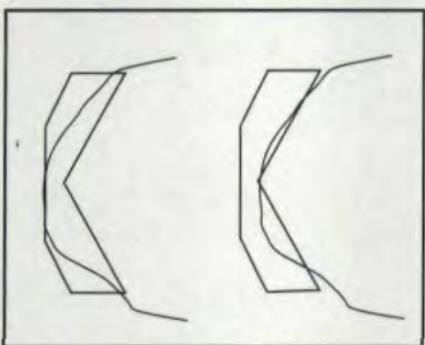

Fig. 11 — "Hollow" is the reversal of curvature.

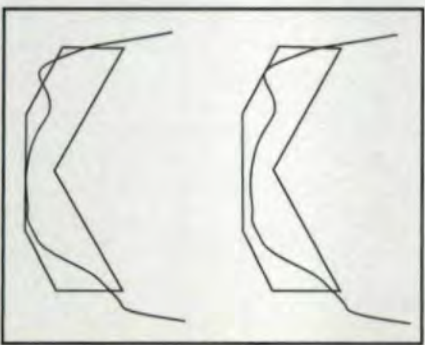

Fig. 12 — A basic K-chart evaluation.

Fig. 13 — A K-chart with more than three points.

and some use the K-chart technique. Most people use some kind of combination of these methods.

Crown of the Surface. There are various computerized techniques available for crown evaluation. Some use a best-fit curve (Fig. 10a). Some compare both ends of the tracing with a high point (Fig. 10b), and some compare a line connecting tracing ends with the highest point (Fig. 10c).

Hollow. Hollow is the reversal of the curvature, as shown in Fig. 11. One can look at hollow as a variation of the form error. This characteristic is widely monitored for evaluation in the automotive industry.

The K-Chart Method of Profile Evaluation

The K-chart is probably the most widely used technique for qualifying or disqualifying a gear. The K-chart is a simple appraisal for deciding whether or not the gear profile is within the specification. However, it is important to note that the K-chart is not a good tool for analyzing the source of a problem. It is only a go/no-go gage which tells whether the profile is good or bad. When a gear does not fit the K-chart, a more detailed analysis must be conducted in order to find and eliminate the source of the problem; e.g., total error must be broken down into the slope and form components.

Despite its seeming simplicity, the K-chart can become a matter of controversy. Many companies—and sometimes different people within the same company—differ about how to interpret a K-chart. For example, is the tracing in Fig. 12 inside or outside the K-chart tolerance?

The answer depends on which interpretation method is used. An inspection tracing could be justified to a "plus" material condition as shown on the left of the figure or a "minus" as shown on the right. Sometimes, K-chart bands are defined with more than three points. This opens up a further proliferation of evaluations. Some people may justify the high point located anywhere between SAP and EAP, as shown on the right of Fig. 13. But some may use a specific range of roll angle for justifying the high point. An example is shown in Fig. 13 on the left, where a middle portion of the tracing is used for justifying a high point of the involute. As a result, the same tracing could be considered as outside (left) or inside (right) a K-chart.

Conclusion

Basic principles of gear inspection have not changed during the last 30 to 40 years. But there has been a dramatic proliferation of gear inspection standards, evaluation techniques and inspection machines. Computers certainly have contributed a great deal to this proliferation.

While the proliferation of gear inspection machines was a welcome sign for gear manufacturers, the current variety of home-grown gear evaluation standards and techniques have had both positive and negative effects. On one hand, it has opened up choices and provided fertile soil for creativity. But on the other hand, especially for people without a strong background in gear geometry, this proliferation of standards and techniques has become a very confusing matter.

AGMA standards for involute and lead evaluation are by no means comprehensive and conclusive. For example, AGMA does not classify form and slope error components. Perhaps that is one of the reasons why many American gear manufacturers have created their own, more detailed, but frequently contradicting standards and techniques. Examples of these contradictions are the aforementioned K-chart and crown varieties of evaluations. These varieties are wide open for different interpretations and resulting disagreements.

To avoid these disagreements, it is helpful to recognize these varieties and, if necessary, develop or adapt, document and communicate the company's policy regarding profile evaluation methods and interpretations of the K-chart or crown.

In contrast, the European gear inspection standards are more comprehensive and adopted widely. As a result, these standards are much more effective in helping gear companies find common ground when dealing with one another. These standards can also be more effective for debugging gear manufacturing processes, for example, by applying slope and form error components.

Regardless of where one stands on the merits of home-grown standard and evaluation proliferation, understanding the basics can help one navigate in this sea of inspection standards and evaluation techniques. ○

Acknowledgement: *The article is based on paper presented at the SME Gearing Conference, Feb. 3, 1997, Chicago, IL.*

Yefim Kotlyar
is the Gear Technology and Processing Manager at Bodine Electric Company, Chicago, IL. He is also the author of a number of articles on gear-related subjects.

Tell Us What You Think . . .
If you found this article of interest and/or useful, please circle **212.**