The Mathematics of Datums†

Theodore H. Hopp
National Institute of Standards and Technology.

If I want to check whether a part is good, I measure its critical dimensions and compare my measurements to the part tolerance. But do I really know exactly what a part tolerance means? What do the numbers I get from measurement have to do with numbers on the part drawing? These questions are the focal issues of the Y14.5.1 Working Group on Mathematical Definitions of Dimensions and Tolerances. This article highlights some of the more challenging technical issues facing this group.

The Y14.5.1 Working Group (originally the Y14 *ad hoc* Committee on Mathematization) has organized its work around the tolerance types defined in ANSI Y14.5, the American National Standard on Dimensioning and Tolerancing. The Group is analyzing how each tolerance type (size, profile, location, etc.) applies to different part geometries. The Group decided early on to treat the formal definition of dimensions as part of the definition of tolerances.

The most difficult problem facing the Group is not so much developing formal mathematics for tolerances, but developing formal mathematics that match engineering common sense. One example of this has to do with the definition of a datum. It is an unfortunate fact of life that physical part features, including datum features, are not perfect. The Y14.5 standard, recognizing this, has the following language:

> If irregularities on the surface of a primary or secondary datum feature are such that the part is unstable (that is, it wobbles) when brought into contact with the corresponding surface of a fixture, the part may be adjusted to an optimum position, if necessary, to simulate the datum.

This statement is lacking in two critical areas, both of which have inspired what I would call highly motivated discussions at committee meetings. First, what is *wobble*? The nonsense we have to avoid is pushing the part up on an edge and calling that wobble. Secondly, what is an *optimum position*? When is adjustment *necessary*? Although many people have a strong intuitive sense of what to do, there is little guidance available from standards or from technical literature. The Group has struggled long and hard to define datums. The challenge is to do so in a way that includes what we want to include and yet cuts out nonsense. While we have not yet fully succeeded, we have made considerable progress. Perhaps presenting our thinking will stimulate a reader of this article to improve on our current solution. An improvement would certainly be welcome.

Finding an Optimum Position

As a way of analyzing the problem, the Group began with a list of criteria for what is a good definition. We also assigned weights to the criteria, from zero meaning not important to ten meaning a "must." The criteria, and the weights we assigned them, were:

1. Does the definition conform to Y14.5? (weight: 10)
2. Does the definition yield a unique solution? (weight: 6)
3. Is the definition mathematically unambiguous? (weight: 10)
4. Is the definition measurable? (weight: 0)
5. Does the definition convey design intent? (weight: 10)

I should explain a few points about this list. Since we have been complaining that Y14.5 does

†Reprinted from **ASPE Newsletter**, September 1990, American Society for Precision Engineering

not define datums, what does criterion 1 mean? The answer is that Y14.5 has an extensive treatment of both the philosophy and practical use of datums. We felt it important to be consistent with the principles laid out in Y14.5. On the other hand, we felt that Y14.5 was not complete regarding overall principles. Hence we added criterion 5. An ideal dimensioning and tolerancing standard would make criteria 1 and 5 synonymous. Finally, the weight assigned to criterion 4 has been somewhat controversial. Our reasoning was that, while measurability is very important in practice, we are trying to define the ideal to which measurement practice should strive. The ideal should be independent of how well we can actually do. If our definitions are so impractical as to be useless as an ideal, the problem really is more basic. The conflict will be between the criteria to which we are faithful (Y14.5 principles, mathematical clarity, and expression of design intent) and practical requirements. We believe there is no basic conflict. By including measurability as a criterion, however, we have stimulated useful discussion of possible definitions.

The Group discussed seven possible ways of associating a primary datum plane with a datum surface on a part. We then rated the definitions against the criteria, assigning a score between zero (violated the criterion) to ten (conformed to the criterion). I will first state and discuss each candidate definition, then I will discuss the results of our evaluation and what we finally resolved. Our result was somewhat unexpected (to us). While it is not entirely satisfactory, it overcomes major weaknesses of the other approaches.

The results of our early discussion are shown in Table 1. The scores reflect the range of opinions of the Group. I will define each method we considered and describe its strengths and weaknesses. (Unless otherwise stated, all definitions are for all the points on the datum feature.)

Position on a level surface in a gravitational field. We imagine the surface to be an infinitesimal sheet and allow it to rest on a datum plane. The stable positions of the sheet determine the allowed datum plane orientations.

The idea is to approximate placing the part on a surface plate and letting it rest freely. We framed the definition using an infinitesimal sheet because we did not want to be concerned with unbalanced parts. That we felt to be in the same category as free-state variation: if it is an issue, the designer must specify how forces are to be applied to the part before establishing a datum.

This definition appeared at first to be our best one. The major weaknesses were that it did not allow "equalizing the rock" and it did not yield a unique solution. (Imagine a ∨-shaped surface.) There was some ambiguity about setting the direction of the gravitational field, but we felt this would not be a serious problem.

Least-squares plane translated to a tangent point. The datum plane is defined by a two-step process. First find the orthogonal-distance regression plane fit to all points on the surface. Then find the plane parallel to this fit but just tangent to the part surface.

The only strength of the least-squares definition is uniqueness. It will almost always result in a part being tilted up in the air about the datum.

Minimax for all points. The idea is to equalize the gaps between the datum and the datum feature: pick the orientation that minimizes the largest gap.

This approach is advocated by the ISO Technical Committee 10 in Technical Report ISO/TR 5460-1985. A strong argument against this definition is that it depends on feature points that will never affect an assembly. Why should the datum change if the valleys of the feature change? The valleys will not constrain assembly or affect function.

	10 Conformance to ANSI Y14.5	6 Yields a unique solution	10 Mathematically unambiguous	0 Measurable	10 Conveys the designer’s intent
Position on a level surface in a gravitational field	10-10	3-3	10-7		10-6
Least-squares plane translated to a tangent point	5-0	10-10	10-10		5-2
Minimax for all points	5-0	8-4	10-8		5-2
Minimum rock	3-0	10-10	10-6		3-0
Minimax for all high points	10-4	8-4	10-8		8-4
All simultaneous rocked positions	2-0	0	10-6		6-0
Single arbitrary rocked position					

Table 1. Summary evaluation of approaches to defining a primary planar datum

Minimum rock. This definition approximates "equalize the rock." The idea is to orient the part so the maximum angular tilt is minimized.

As with the least-squares tangent plane, this definition can force orientations that do not conform with design intent. On the strength of previous conversations with Y14.5 committee members, we also felt that this definition was at odds with the principles of the Y14.5 standard.

Minimax for all high points. This is an attempt to address the weakness of the minimax-for-all-points definition. The "high points" are all points on the surface that can be contacted by a plane. The datum plane is chosen to minimizes the maximum gap to the high points.

The Group felt that this definition was generally sound. However, it still suffered from the problem of forcing nonsensical datum orientations. The problems were not as severe as with other definitions.

All simultaneous rocked positions. This definition says that all rocked positions are datums *simultaneously*. The part must satisfy all tolerances for every possible orientation of the datum plane.

This definition is based on the following heuristic argument. When a part is assembled, the orientation of it’s datum surfaces will become fixed. Any one of the possible orientations is as likely as another. Therefore, the part should meet functional requirements for any possible orientation. The Group felt that this constraint is too strict.

Single arbitrary rocked position. This is the converse of the previous definition. The orientation that makes the part look the best defines the datum.

The heuristic argument is that, if the part does not work properly at one orientation but does at another, it can easily be placed into the orientation that works. Perhaps because of the word *arbitrary*, the Group felt that this definition was too fast-and-loose. It was ruled out *a priori* and not even rated.

Armed with the results of our early brainstorming, we visited the Y14.5 Committee to ask their views. Here is where the power of our research-by-committee approach had an opportunity to shine. Unfortunately, our committee thinking had been somewhat dull. The Y14.5 Committee, in no uncertain terms, told us that the only definition that made engineering sense was the one we had decided to ignore. (I might add that quite a few engineers were a party to our decision.)

On reflection, the guidance of the Y14.5 Committee seems closest to what designers, manufacturers, and inspectors use in practice. Our original fixation on defining a single, optimum datum for a feature blinded us to an important point: a datum does not exist for itself; it exists in relation to, and for the sake of, features that are controlled through a datum reference system. The "single arbitrary rocked position" definition captures exactly this idea.

The Y14.5.1 Group formalized datums using an idea we called *candidate datum sets*. These sets are defined as all the allowed rocked positions of the datum feature about the datum. We are now working out the details of when and how to select a particular datum from the candidate datum set; in particular, how to deal with simultaneous and separate requirements. This formalization led us to a useful insight regarding rocking datum features and datum features of size. A datum feature of size, produced away from the maximum material condition (MMC), and referenced at MMC, allows the datum to shift on the part. This shifting can also be handled by defining candidate datum sets. Mathematically, the only distinction between rocking datums and datums of size (which, by the way, can also wobble, as well as depart from MMC) is the mechanism that gives rise to more than one candidate datum.

Defining Wobble

We had one task left—to define what all the rocked positions are. To date, this is still our major weakness. All the problems noted in the definitions we considered came back to haunt us when we tried to define the limits of rock. We have not developed a definition that includes all the positions that engineering common-sense would allow, yet excludes tilting the part up in the air in unacceptable ways.

The crux of the problem can be seen in Figure 1. Here, we show the evolution of a datum feature from a surface that no one would consider a rocker to a surface that everyone believes is clearly a rocker. What mathematics allows the bottom case to rock onto the bevels, yet does not allow the part at the top to rock onto a chamfer?

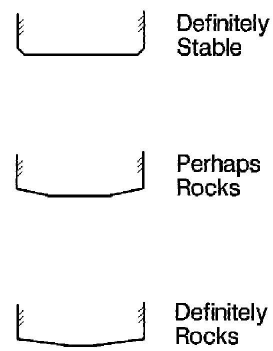

Figure 1. The ambiguity of *all rocked positions*. As a datum feature degrades in quality (from top to bottom) when does it become a "rocker"?

Tradition has it that rocking is controlled by controlling form: a flatness tolerance will limit rocking of a planar feature. While this may be true indirectly, it is not strong enough. By that prescription, a (hypothetical) perfect datum feature can then be tilted into the air by as much as the flatness tolerance, to allow a controlled feature to come into tolerance. We were back to the nonsense we wanted to avoid.

We have also discovered one other difficulty. We now have a situation in which a datum can be purposely degraded (e.g., by hitting it with a hammer) simply to introduce wobble. The wobble can then be used to make the part look good. We are all very uncomfortable with the idea that the part can be made to look better by making one feature worse. Unfortunately, this difficulty seems to be innate in the idea of optimizing a rocking datum and not one we will be able to overcome with clever mathematics.

With the above issues in mind, we began the work of defining candidate datum sets. We wanted the definition to reflect the actual quality of the datum feature, rather than the design tolerances. Tom Charlton (of Brown & Sharpe), noting that we had not found it difficult to define candidate datum sets for features of size, suggested that we needed a concept of size that could be applied to planar features. The actual flatness of the feature seemed to be the measure of size we would need. Our current definition of candidate datum sets is based on flatness.

The idea is to use the actual flatness zone of the datum feature to control the orientation of the datum. The definition is illustrated (in cross-section) in Figure 2. The orientation, width, and extent of the flatness zone determine the extremes of how the datum can be tilted about the datum feature. The actual datum is shifted to be tangent to the feature.

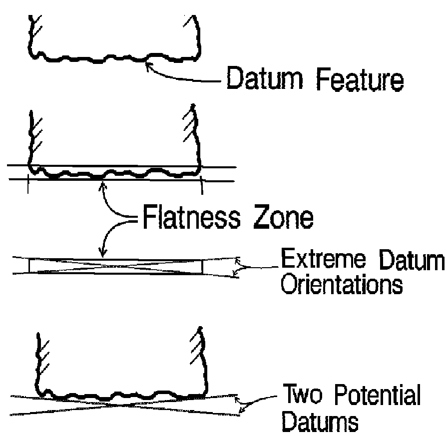

Figure 2. An illustration of how flatness defines the candidate datum set.

This definition has two strengths. First, it eliminates all nonsense such as having the feature tilted up on edge about the datum. As a datum feature gets better, the set of candidate datums is reduced; a perfect feature has exactly one candidate datum. Second, the definition does allow the "datum leveling" practices that are commonly used.

We are not happy with the definition, however. The reason is that it does not allow *all* the positions we would like to allow. Consider, for example, a datum feature manufactured in a "∨" shape. A few minutes consideration will show that our definition will force the datum feature to be balanced on the point of the ∨. The part cannot rest flat on one side or the other, as it might on a surface plate or in an assembly. While we are not content with our current definition, we feel we have come very close to capturing in mathematics, the engineering common sense that we have used as our model.

Summary

Datums have been one of the most difficult problems the Group has dealt with. With the definition outlined above, we feel we now have the problem under control. While our definition is not perfect, we can live with it. We are now in the process of expanding this definition to cover all aspects of datum use described in the Y14.5 standard.

Other difficulties still face the Group, including potentially challenging problems having to do with size and with boundaries of features. The Group will meet again in September near Washington, D.C. I am confident that, as at past meetings, we will continue to make good progress in formalizing dimensions and tolerances. Meetings are open to anyone who wants to contribute to this effort. If the above history of our struggle with datums has piqued the interest of others, I can think of no better place to pursue that interest than at one of our meetings.