Self-Consistency and Transitivity in Self-Calibration Procedures

by

Michael Raugh

Technical Report: CSL-TR-91-484

July 1991

Computer Systems Laboratory

Departments of Electrical Engineering and Computer Science

Stanford University

Stanford, California 94305-4055

Abstract

Self-calibration refers to the use of an uncalibrated measuring instrument and an uncalibrated object called an artifact, such as a rigid marked plate, to simultaneously measure the artifact and calibrate the instrument. Typically, the artifact is measured in more than one position, and the required information is derived from comparisons of the various measurements. The problems of self-calibration are surprisingly subtle. This paper develops concepts and vocabulary for dealing with such problems in one and two dimensions and uses simple (non-optimal) measurement procedures to reveal the underlying principles. The approach in two dimensions is mathematically constructive: procedures are described for measuring an uncalibrated artifact in several stages, involving progressive transformations of the instrument's uncalibrated coordinate system, until correct coordinates for the artifact are obtained and calibration of the instrument is achieved. *Self-consistency* and *transitivity,* as defined within, emerge as key concepts. It is shown that self-consistency and transitivity are necessary conditions for self-calibration. Consequently, in general, it is impossible to calibrate a two-dimensional measuring instrument by simply rotating and measuring a calibration plate about a fixed center.

Key Words and Phrases: ball plates, calibration, Coordinate Measuring Machines, measurement artifacts, self-calibration procedures, pattern accuracy, precision engineering, precision metrology, self-consistency, transitivity

Contents

Figures

Introduction

The purpose of this paper is to develop concepts for a theory of two-dimensional *self-calibration* and to derive some basic principles stating conditions under which self-calibration is possible.[1] Self-calibration procedures are described, but the procedures are not optimized. Emphasis is placed on developing straightforward procedures that can be analyzed easily to reveal fundamental issues of self-calibration and to demonstrate fundamental principles. The problems of self-calibration are surprisingly subtle, and I have found it convenient to introduce a special vocabulary for dealing with them. The glossary, therefore, is an important part of the paper. To make the ideas of the paper accessible in more than one way, I have made the figures and the glossary somewhat self-contained. The principal sections are:

- Preliminary Examples and Definitions
- Self-Calibration Illustrated in One Dimension
- Self-Calibration in Two Dimensions
- Summary, Conclusions, and Exercises
- Glossary

Preliminary Examples and Definitions begins by defining calibration and discussing the dual relationship of *calibration* and *measurement.* Three examples of calibration and measurement are given to contrast *standard calibration problems* with self-calibration problems and to illustrate the concept of *transitivity.* At least on the face of it, the two kinds of problems appear to be very different. The concept of *self-consisfency* is introduced. *Ball plates* and *Coordinate Measuring Machines (CMM),* well-known devices in metrology, are used to illustrate two-dimensional problems. They are described briefly, and references are supplied for readers who are unfamiliar with them. Axioms regarding coordinate systems are formulated, and *coordinate patches* are defined.

Self-Calibration Illustrated in One Dimension presents more new vocabulary in the simplest context, namely, on the line and on the circumference of a circle. Simple examples of the use of self-consistency and transitivity are given. The concepts of *determinacy,* and *indeterminacy* are introduced.

Self-Calibration in Two Dimensions, the main section of the paper, uses examples of ball plates and Coordinate Measuring Machines to develop a theory of two-dimensional calibration. The problem posed is whether it is

---

[1] Italics are used to introduce terms that are either new or of special importance in the exposition. Most of the italicized terms are defined in the text and in the glossary and may also be defined in the figure captions.

possible to calibrate a two-dimensional measuring instrument by measuring a rigid pattern of points in various placements on the measuring plane of the instrument, if the pattern is not known or is known imprecisely. The main result is **a** formulation of the fundamental *principle of self-calibration,* which states that self-calibration is possible only if the measurement procedures are *transitive.* Constructive procedures are explained for deriving a calibration from appropriate self-calibration procedures.

The Summary, Conclusions, and Exercises summarizes the results of the paper, suggests some future directions, as, for example, calibration in three dimensions, and proposes exercises that extend the theory. It also describes briefly some connections with Raugh (1985).

The Glossary gathers into one place the terms that have been italicized in the text. Many of the words have been given new meanings, which may cause difficulty on first reading. To alleviate confusion, the glossary contains definitions and cross-references so that the reader may see how the new terms are interrelated. Many of the more important terms are repeatedly defined in the text or footnotes and in the figure captions.

Preliminary Examples and Definitions

To calibrate means to assign physically meaningful and correct numerical values or coordinates to the graduations of a measuring device, as for example on a ruler or a thermometer. As viewed in this paper, calibration goes hand-in-hand with measurement -- they are dual operations. Ordinarily, measurement standards are used to calibrate a measuring device. Conversely, calibrated measuring devices are used to measure objects. Thus, measurement and calibration problems are chicken-or-egg. This is what makes the problem of self-calibration, i.e., the problem of using an uncalibrated measuring device to measure an uncalibrated object, so interesting and challenging.

When we speak of measuring an object, such as a pattern of grid points or VLSI-circuit features, we mean either finding coordinates for points of the object in a well-defined coordinate system or else determining the *shape* of the object by some other means, as for example measuring the angles, or the lengths of the legs, of all the triangles that can be drawn between any three points of the object. For example, measurement of a two-dimensional pattern of grid points can mean either determining all the *internal angles* of the object or determining coordinates for the grid points in an *orthonormal* *coordinate system.*[2] To measure an object accurately, ordinarily a calibrated measuring device must be used.

Conversely, when we speak of calibrating a measuring instrument, we have in mind a device that can assign coordinates to the points of a measured object, but the coordinate system of the device is systematically inaccurate. Usually the measuring instrument will contain some unknown curvilinear component of distortion.[3] For example, in two-dimensional metrology using, say, a metrology tool or an electron beam lithography system, laser interferometers are used to track the motion of a measurement stage on which the object to be measured is mounted, or to track a probe that moves over the object from point to point. But, because of the typical imperfections of machinery, such tracking mechanisms do not yield precisely orthonormal coordinates. To underscore the fact that raw instrument coordinates contain systematic error, I sometimes refer to them as *coordinate markers.* They are akin to the graduations of an uncalibrated thermometer or measuring stick, and they need to be calibrated. So, by calibration of the instrument, I mean a process of determining a transformation of the instrument's coordinate markers to a well-characterized coordinate system, in effect eliminating the distortion. In this paper, such a corrective transformation is called a *calibration function,* or *calibration mapping* (or, simply, calibration); here it is assumed that the effect of calibration is to transform an instrument's coordinates into an orthonormal coordinate system. (See Figure 1.) To calibrate an instrument, we need to compare the instrument's coordinate system with an accurately measured object.

There are many approaches to calibration, which superficially seem to be quite different but on deeper analysis are all similar. Three examples follow. The first is actually an example of measurement that highlights the intimate relationship between measurement and calibration. The second example illustrates a classical calibration problem, which I refer to as the standard problem. The first two examples illustrate transitivity, which, conceptually, involves the direct or indirect comparison of all lengths and angles of an object to one another. More precise definitions will emerge later in the paper. The third example illustrates the problem of self-calibration. A theme developed in the paper is that, even in self-calibration, the all-

---

[2]By orthonormal coordinate system, I mean the kind of coordinate system used in classical analytic geometry familiar to anyone who has studied calculus, namely the *Cartesian* method for handling euclidean geometry in a systematically algebraic and numerical way. See Glossary.

[3]In other words, if a straight line, or grid of equally-spaced straight lines, half of which are oriented at right angles with respect to the other half, is measured on the measuring instrument, and the coordinate markers are plotted out on orthonormal graph paper, then some or all of the straight lines will be plotted as curved lines, and the curves will no longer necessarily be equally spaced or cross each other at right angles.

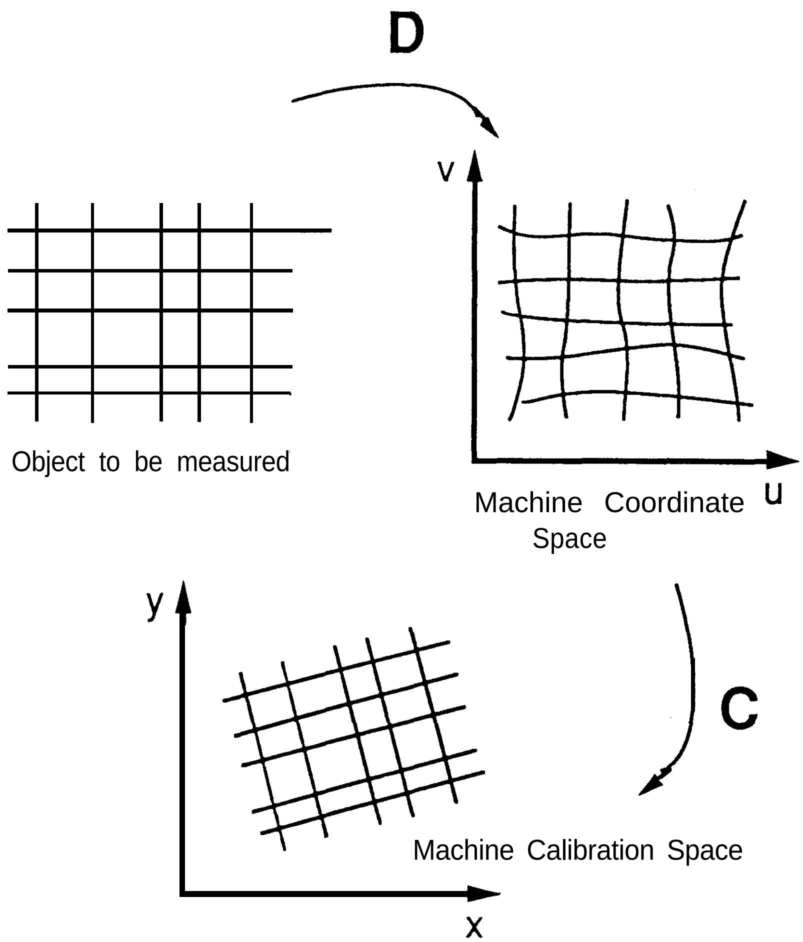

Figure *1. Calibration interpreted as a mapping.* Two coordinate systems are used to illustrate the problem of calibrating a measuring instrument. An uncalibrated object to be measured by an uncalibrated instrument is shown at the upper left. The *(u , v)*-coordinate system at the upper right refers to actual machine measurements, i.e., the coordinates that the machine itself assigns to the points of the measurement plane. When an object such as the rectangular grid on the upper left is measured and its image plotted on orthonormal graph paper as shown in the Machine Coordinate Space at the upper right, the image of the object is distorted by the non-orthogonality (or curvilinearity) of the instrument's coordinate system. The distorted measurements are represented symbolically by the mapping D. The plotted image is all you can see -- you cannot see the true shape of the object. The problem of calibration is to find a mapping C that transforms the instrument coordinates (u, *v)* into coordinates, say (x, *y),* which, when plotted on orthonormal graph paper, yield a non-distorted image of the object, as in the Machine Calibration Space at the lower center. Calibration, in the sense employed here, allows that the size, position, and orientation of the image at the lower center may differ from the original object, but the shape must be the same.

embracing self-comparison, i.e., transitivity, is a necessary feature of valid calibration procedures.

First Example of Calibration: Topographical Surveys. Surveyors measure a topographic domain using calibrated instruments when they determine accurate coordinates, such as longitude, latitude, and elevation, for a subset of the natural features of the land or for an array of benchmarks on the land. By forming a chain of interlocked triangles and determining coordinates for each vertex throughout a domain, they can reference all points of interest to known vertices. The land thus surveyed becomes a sort of calibrated instrument in the sense that future local surveys can use the original benchmarks for interpolating coordinates for additional features of the land. The process of surveying is one of derivation of coordinates from triangulation measurements. It is important to note that in this process, the surveyor carries with him surveying tools to serve as accurate standards of length and angle against which to measure local features. For example, the surveyor may use a theodolite to measure vertical and horizontal angles and a laser-ranging instrument to measure distances, or a tachymeter, which performs all of these functions in one device. Thus an angle or length at one station on the land can be compared with an angle or length at another station through the mediation of such portable instruments. When multiple instruments are used by cooperating surveying parties, even the transits, theodolites, and tachymeters themselves must be compared, either one to another, respectively, or each to an appropriate standard, to ensure mutual consistency. The instruments, in other words, must be calibrated. This process of comparison through the mediation of calibrated instruments makes use of the logical *principle of transitivity,* namely, the fact that "two quantities equal to a third are equal to each other." The example shows how each measurement depends upon the use of some previously calibrated tool. Through additional examples, we shall see that implicit or explicit mutual comparison is universal in calibration problems.

Second Example of Calibration: Standard Calibration. A measuring instrument is known to distort measurements and we wish to calibrate it. A *measurement standard,* i.e. an accurately measured grid plate or other artifact, is used such that, for example, each point of the grid is referenced to an orthonormal coordinate system with a degree of accuracy exceeding the reproducibility of the instrument. The measurement standard may have been obtained from the National Institute of Standards and Technologies (NIST). How it was made is of no concern; we just need to know that the given coordinates are accurate. Suppose that the standard grid plate is fastened to the *measuring plane* of the measuring instrument, and a least-squares method is used to fit a mapping function of the measurements to the known coordinates of the grid. The function determined in this way is a calibration mapping. In this example the grid itself serves to convey a measurement standard to each local region of the instrument's measurement domain. In a sense, the grid does for the instrument what the surveyor and his portable measuring instruments do for uncharted territory. The analogy goes deeper inasmuch as the standard grid was probably made and measured by a process strictly comparable to the surveyor's method, using tools such **as** rulers and protractors (or more precise equivalents) that were moved through the "territory" of the grid to determine the angular and length relationships among the grid points. Thus, the principle of transitivity is involved in the mutual comparisons of lengths and angles, respectively, that are implicit in the measurement process.

Third Example of Calibration: Self-Calibration. We want to calibrate an instrument for which there is neither a convenient measurement standard nor a convenient means of conveying a standard to each local portion of the instrument's measurement domain. This problem exists, in particular, in high-precision lithography, wherein VLSI circuits are printed on a scale of precision for which no correspondingly accurate standard grid can be manufactured at this time. The problem seems to be quite different from the first two examples. It appears that the only means of testing measurement accuracy for this kind of problem is to measure a rigid, complex *pattern,* such as an uncalibrated grid, in more than one position on the measuring plane, and to compare the results. Hopefully, we can deduce both a calibration for the instrument and accurate determination of Cartesian coordinates for the rigid pattern directly from the measurements. This approach to calibration is called self-calibration because the instrument is used to calibrate itself. Actually, it is more correct to say that the instrument and the unmeasured pattern are used together to calibrate and measure each other simultaneously, illuminating the duality of calibration and measurement. It is this kind of problem that concerns us in this paper.

The third example of calibration, self-calibration, appears to be different from the second example of calibration because it does not use an accurately measured standard to test the performance of the instrument to be calibrated. Nevertheless, the multiple measurements of the pattern used in the third example can be compared to one another. In fact, in the absence of a measurement standard, these self-comparisons can serve to make the pattern itself a substitute for a measurement standard. This gives rise to the notion of *self-consistency,* as defined below. It will turn out that self-consistency and transitivity are the key concepts in self-calibration.

Self-Consistency and Machine Coordinate Space. The **idea** of self-consistency arises in the following way. Suppose that we have a calibrated two-dimensional measuring instrument that gives correct *Cartesian coordinates* for measured points. Suppose, then, that we use the instrument to measure a distinctive pattern of points on the surface of a rigid plate. For example, let the points be arranged in the shape of a cat. If we plotted each of the measured points on orthonormal graph paper, we should expect to see an undistorted image of the cat, although the image may be of a different scale, either larger or smaller than the original. Furthermore, if we moved the rigid plate to another position on the instrument's measuring plane, remeasured it, and plotted the new set of points, we should expect to see a second image of the cat on the graph paper, identical in size and shape to the first image but in a different position. Likewise, if we continued to move the rigid plate to various positions on the measuring plane, remeasured the points and plotted them, we should see various identical images of the cat in correspondingly various positions. The point of all this is that, if the measuring instrument is calibrated, then various images of the same object should all be *congruent* to each other. In other words, all the images of the object would be of identical size and shape. Self-consistency is the term I use to say that a specific set of images of a measured artifact are all congruent to one another, without regard to the shape of the original pattern of points. Thus self-consistency is a necessary property of a calibrated measuring instrument, so it is natural to ask whether self-consistency among a specific set of images of a measured object is also sufficient to determine that the measuring instrument is calibrated. This is the key question of the paper. Before taking it up, I want to discuss self-consistency in more detail.

Note that the term self-consistency is used with respect to a combination of factors, namely, 1) a specific measuring instrument and its coordinate system, 2) a specific artifact to be measured by the instrument, 3) specific procedures for measuring the artifact in a finite number of positions on the instruments measuring plane, and 3) the resulting measurements and plots of the measurements on Cartesian graph paper. For example, suppose that the artifact is a pattern of points spread out randomly on a rigid grid plate. First, let the grid be fastened in some position on the instrument's measuring table and the points of the pattern be measured in a specified order. Let the coordinates of the machine be denoted by (u,v), and denote the actual measurements for the n points by (u_i, v_i), where i runs from 1 to n. The measurements thus obtained can be thought of as the *image* of the pattern in the *machine coordinate space* of the measuring instrument. The image can actually be depicted on orthonormal graph paper, as follows, although in general the result will be a distortion of the true shape of the rigid pattern if the measuring instrument is uncalibrated. Let the coordinate axes of the graph paper also be designated by (u,v). For each measured point (u_i, v_i), plot the graph-point (u_i, v_i). Second, remove the rigid plate and reposition it in some approximately known position on the measuring table. Now, once again measure the pattern points in the same specific order as before, then project the image of the pattern onto the same graph paper exactly as before. This same process of repositioning the pattern plate, remeasuring the pattern points, and projecting the points onto the (u,v)-image plane, can be repeated as often as is practicable. In general, the

A

resulting pattern images will all differ in size and shape from the original pattern. And, of course, their orientations will be as varied as were the orientations of the rigid pattern in its various placements on the measuring table. But the deciding question is whether all of the pattern images are of the same size and shape as each other, i.e., whether they are all congruent to each other. If so, I say that the machine's (u,v) coordinate system is *self-consistent with respect to the given set of measurements.* Remember, we are asking only whether the images are congruent to one another, not whether they are congruent to the original pattern!

It is important to note that self-consistency is an observable phenomenon, whether or not the measuring instrument is calibrated. We only have to compare the plotted pattern images with each other, without regard to the original pattern. As noted above, if the measuring instrument is calibrated, then any set of measurements will be self-consistent. But if we are using an uncalibrated measuring instrument, we cannot observe whether the plotted images are of the same shape as the original pattern because the instrument gives distorted coordinates. Therefore, a useful theory of self-calibration must enable us to know under what conditions a self-consistent coordinate system may be known to project images of correct shape, i.e., to know whether the coordinate system is calibrated. (See Figure 2.)

To restate the definition, self-consistency of the coordinate system (with respect to a set of measurements) means that the images of the pattern in the machine coordinate space of the instrument all have the same shape (when plotted on orthonormal graph paper), though indeed they may be oriented at a variety of positions corresponding to the variety of positions of the pattern on the measuring plane. Note that a calibrated coordinate system must exhibit self-consistency not only with respect to the given pattern and its various placements in the measurement plane but for arbitrary patterns in all conceivable placements. But the converse is not necessarily true. As we shall see, it is possible for an uncalibrated system to project a set of self-consistent images.

Self-Consistency and Machine **Calibration Space. We** need to take the argument one step farther. Suppose that we have an uncalibrated measuring instrument, an uncalibrated artifact to be measured by the instrument, and a series of placements and measurements of the artifact on the measuring plane, as described above. Moreover, suppose as above that each set of measurements of the artifact are imaged on Cartesian graph paper, but the resulting images are not self-consistent. In this case, we might be able to find a transformation C of the machine coordinates, in other words a mapping from the $(u,$ &coordinate system to another Cartesian coordinate system (x, y) such that the images in the (x,y) system are self-consistent. I call this new coordinate space *machine calibration space.* (See Figure 1 again.) The crucial question is to determine whether there are conditions under which self-

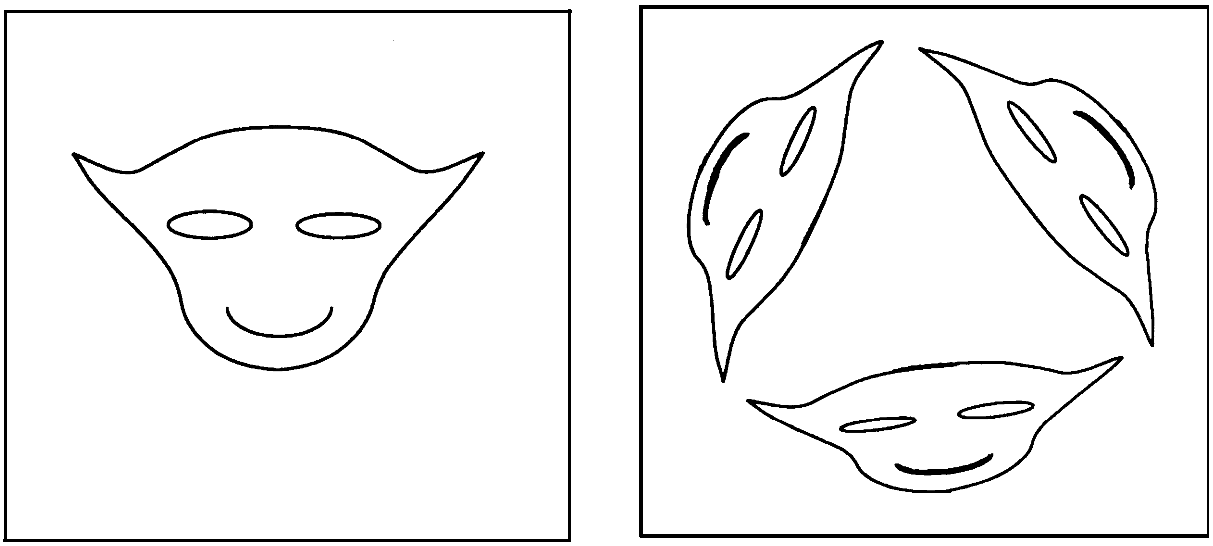

Figure 2. *Self-consistency illustrated by cats.* The pattern of points on the left, in the figure of a cat, is measured by a measuring instrument in three separate placements on the instrument's measuring plane. Each set of measurements is plotted on Cartesian graph paper as illustrated, respectively, on the right, resulting in three images that are located in different positions but are all identical in size and shape. That is to say, all the images are congruent to each other. When this situation arises for a given series of measurements of an object in various positions, in which the resulting images are all congruent to one another, the measurements are said to be self-consistent. It is important to note that if the measuring instrument is calibrated, then the various images on the right that are all congruent to one another will also be of similar shape to the original object on the left but may be of different size. However, if the measuring instrument is uncalibrated, then the images may be of different shape than the original object, as in the illustration. The problem of self-calibration boils down to determining conditions under which self-consistency is a sufficient condition for calibration.

consistent mappings in this sense are necessarily correct calibrations, and; if so, whether it is possible to use the raw measurements to derive such a calibration mapping C. The answer is that practical procedures can be implemented to derive a valid calibration, but, as hinted earlier, the procedures must satisfy transitivity criteria that will be developed in later sections.[4]

The Local Linearity of Machine Coordinates. We turn now to consider the way in which a measuring instrument assigns coordinates to points in space. Suppose that we have. a two-dimensional measuring instrument. Flat objects may be placed upon the measuring plane of the instrument, and machine coordinates may be obtained for any point designated on the flat object. The calibration problem arises when the exact nature of the coordinate system is unknown. It has been said that we wish to have orthonormal coordinates but that the machine departs from this ideal in some unknown curvilinear way. Without real loss of generality, I suppose that the *coordinate function,* i.e. the mapping that assigns coordinates to points of the measurement plane, is a suitably smooth, one-to-one invertible mapping of the plane (e.g., continuously differentiable with non-singular jacobian). In that case, the coordinate function may be regarded as linear to first order in a small neighborhood of each point.[5] This local linear function may vary from place to place. Generally, the local linear coordinates will not be orthonormal. For the sake of simplicity, I idealize the situation by supposing that at each point the coordinate system is exactly linear, at least within a small neighborhood of the point, and that there exists a practical means of transforming the local linear coordinates to orthonormal coordinates. In other words, I will assume as axiomatic that each point is contained in a small neighborhood throughout which we can obtain a perfectly orthonormal coordinate system. There is little loss of generality because it is theoretically possible to take a neighborhood small enough so that our calculations will be

[4]Vocabulary adapted from mathematics is introduced throughout the paper. At appropriate locations an explanation for each adapted term is given, accompanied by reference to its mathematical usage, as for example the word transitivity, the meaning for which will be refined in later footnotes.

[5]Note that such a coordinate function, or mapping of points of the plane to points of a coordinate space, implies that the points of the plane may be represented in terms of some "correct" underlying coordinate reference system. Because we generally assume that we operate within a euclidean universe, we may suppose that the underlying reference frame is orthonormal. Note, however, that such a reference frame is a purely theoretical construct. In fact, if we had an explicit representation of the machine coordinates as a transformation of coordinates from such an orthonormal frame to the coordinate space, there would be no need of calibration -- the inverse of the coordinate transformation would provide a calibration function. Now, regarding the coordinate function as such a transformation, we see that the local linear approximation to the coordinate function is given by the *differential* of the transformation (Buck 1965, 263). The linear transformation so obtained is also referred to as the *tangent linear transformation*, and its coefficients are given by the *jacobian matrix* of the transformation (Protter and Morrey 1964, 505-507).

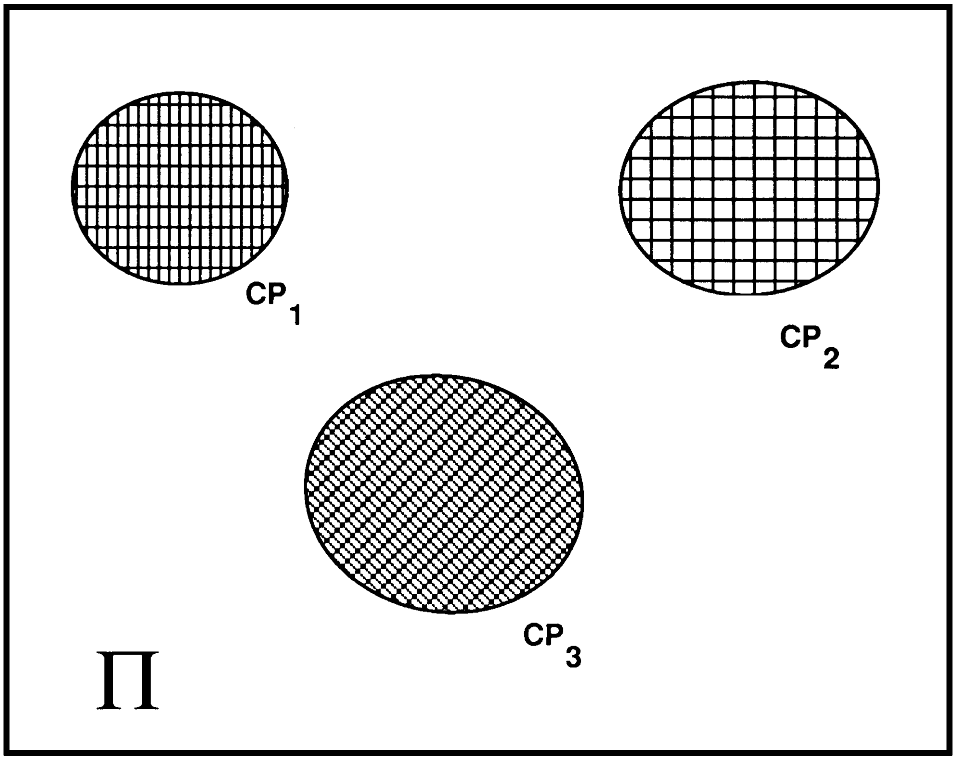

Figure *3. Magnified view of three Coordinate Patches.* Π represents the measurement plane of a two-dimensional measuring instrument. Three local regions are shown, in each of which it is assumed that the machine's coordinate system yields linear coordinates that are not necessarily orthonormal but can be made orthonormal as shown in Figure 3 by a process described in the text. The coordinate system in each such local region is called a Coordinate Patch (CP). It is assumed that the local coordinate systems vary from place to place on the measurement plane in some unknown way. And the relative scales of the various axes, angles of inclination of the various axes to each other, and positions of the origins are unknown. This is a simplified representation of an instrument-coordinate system that is curvilinear in an unknown way.

as accurate as need be. I refer to such a neighborhood as a *coordinate patch* (CP).[6] (See Figure 3.)

Ball Plates. We are now nearly finished with the preliminary examples and definitions and will soon embark on the development of a theory of self-calibration. This is a good place to pause for a moment and reflect on where we are heading. The basic ideas and principles of self-calibration will be developed in a progressive series of examples. We will start in the next section with the simplest examples in one dimension. The examples are trivial, but they allow us to easily illustrate the importance of self-consistency and transitivity. Afterwards, in the following section, we will turn to the more challenging problem of calibration in two dimensions. For the latter case, I had to make a choice. I could have used flat plates with various kinds of patterns of points on them to illustrate the problems. Such an approach would be the most straightforward mathematically. But I decided instead to use the more concrete examples of *ball plates.* The nature and calibration of ball plates, and *Coordinate Measuring Machines (CMM)* used by metrologists to measure ball plates, has been treated in early papers by Reeve (1974) and Hocken & Borchardt (1979). I recommend these papers to readers who are unfamiliar with these devices. I hope that this choice enables the reader to more easily imagine practical procedures in the discussion of self-calibration in two dimensions. Before proceeding, I want to give a brief introduction to ball plates and outline some of the mathematical issues concerning their use in calibration.

A ball plate consists of very accurately turned and well-matched small spheres mounted securely on a dimensionally stable plate. For my purpose, I assume that each ball is small enough to lie entirely within a coordinate patch of the Coordinate Measuring Machine that will be used to measure it. I also assume the following. The ball plate can be mounted on the stage of a Coordinate Measuring Machine, whereupon the spheres can be probed and measured. The multiple measurements taken at the surface of each sphere can then be used to determine and correct the metric properties of the Coordinate Measuring Machine in the locality of the particular sphere. There are subtle problems in using such local measurements to rectify local coordinates. I assume that by probing the circumference of a perfect circle, or the surface of a perfect ball, a nearly perfect orthonormal coordinate system can be derived for a small neighborhood around the ball.[7] This assumption

---

[6] The term coordinate patch is borrowed from the theory of surfaces, where it is used to designate a local coordinate system (Protter and Morrey 1964, 604).

[7] Assuming that the coordinate system is strictly linear in the neighborhood of a point, this could be done in two dimensions by measuring a small number of points on an arc of a circle. As few as five such points can be used to determine a transformation of the local coordinate system to an orthonormal coordinate system. This fact may be deduced from the *polar decomposition theorem*, which states that any linear transformation is equivalent to a rotation

may or may not be practical for real calibration, but I assume it as axiomatic in order to examine the fundamental issues. Furthermore, although a ball plate extends into three-dimensional space, I assume that the elevation is relatively slight enough that the three-dimensional coordinates obtained for the surface of each ball can be projected onto the two-dimensional plane without loss of accuracy. By this assumption, I regard a Coordinate Measuring Machine as essentially a two-dimensional measuring device. (See Figure 4.)

Self-Calibration Illustrated in One Dimension

The Line *L* and Circular Arc C. Our objective is to understand self-calibration in two dimensions. Several key concepts can be illustrated easily in one dimension. Consider the examples of an infinitely long straight line *L* and the circumference C of a circle centered at P. We begin with the observation that there are two basic motions of a rigid object: translation and rotation. Accordingly, we imagine translations of *L* (or subsets of *L*) upon itself and rotations of C (or subsets of *C*) about a center π.

Self-Calibration on *L* and C. There are two distinct approaches to calibration. The classical approach to calibration is to compare the object to be calibrated with a well measured object. The first two examples discussed above in the section Preliminary Examples and Definitions are of this kind. For example, in the case of the straight line, one could imagine successive translations of a standard unit interval, both right and left, beginning at an arbitrary starting point designated as the origin, and successively marking off the positive and negative integer points, respectively. Likewise, one could move a standard interval progressively around the circumference of the circle, marking off the cumulative measure at each step. Equivalently, the standard interval could be held fixed, and the object to be calibrated could be translated (if *L*) or rotated (if *C*) by steps of unit length and marked at each step.

Self-Consistencv and Transitivitv. The second approach to calibration, self-calibration, uses an object to calibrate itself without the aid of a measurement standard. This corresponds to the third example discussed in

---

followed by stretching of the coordinates along a set of mutually orthogonal axes; the stretch factor may differ for each of the axes (Halmos 1974, 169; Segel 1977, 184). The same fact is known as a property of the infinitesimal strain tensor, which can be decomposed into a pure rotation and a pure stretch along orthogonal axes (Sokolnikoff 1956, 21). Similarly, for the three-dimensional case, as few as nine points measured on the surface of a sphere can be sufficient to determine a transformation of a general linear coordinate system to orthonormal coordinates. Thus by probing each ball of a ball plate a small number of times, it is possible to deduce an orthonormal coordinate system for the coordinate patch of each ball plate. Note that there is no canonical orthonormal system for any of the coordinate patches. Scale, orientation, and placement of the origin in each case is essentially arbitrary.

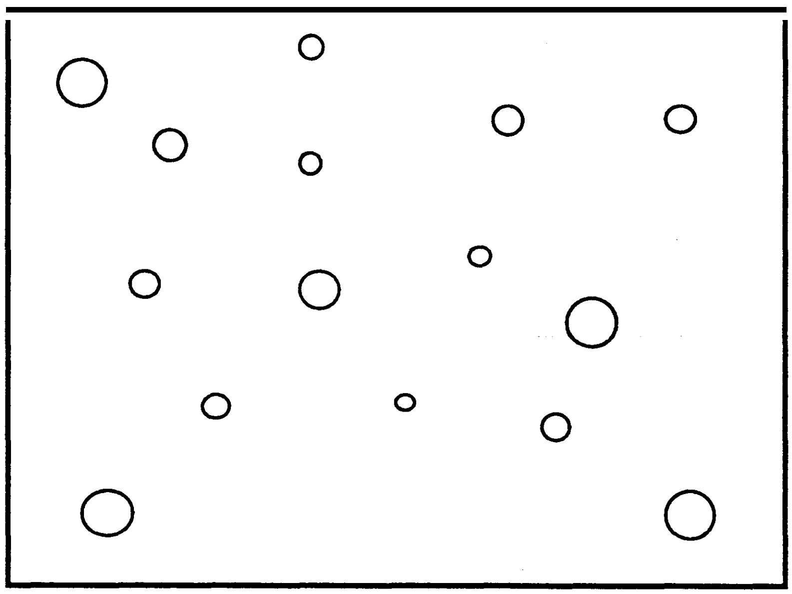

Figure *4. Top view of a random ball plate.* A ball plate is a rigid plate on which small spheres are mounted securely. It is assumed that, when a ball plate is properly fastened on the measurement plane of a Coordinate Measuring Machine, the Coordinate Measuring Machine is capable of probing the exposed surface of each sphere. Furthermore, it is assumed as axiomatic that each sphere is small enough to lie within a coordinate patch of the Coordinate Measuring Machine, and that the measurements of each sphere can be used to *rectify* the linear coordinate system of the local coordinate patch, i.e., to transform the linear coordinates into orthonormal coordinates. We make no claim yet as to the relative scales, orientations, and origins of such derived orthonormal coordinate systems; in fact, the problem of calibration boils down to relating all such local coordinate systems to a common coordinate framework.

the section Preliminary Examples and Definitions. For example, consider the line L, and suppose that L already contains coordinate markings located at intervals known to be at least approximately equal. Suppose that by some means we were able to adjust the markings so that they were self-consistent in the following sense. If a copy of L, call it L', were shifted upon L to the right by exactly the length of the first interval, and it turned out that all the coordinate markings of L and L' exactly coincided, then we would say that the coordination of L is self-consistent with respect to the procedure of shifting by one unit. Here we avoid the question of how to adjust the original markings to achieve self-consistency. We only observe that, if we could achieve self-consistency, then by a simple argument of mathematical induction we could conclude that the intervals are all of an equal length. The idea is that the first interval is compared directly with the second interval, the second is compared directly with the third, the third with the fourth, and so on. Similar comparisons can be made leftward. Thus, by induction and logical transitivity, a sort of cascade of implicit comparisons is made of each interval to every other interval, both left and right of the first interval. The result is that we obtain a pattern of points of verifiable structure. (See Figure 5.)

By the term calibration of a structure, I mean the determination of its geometric shape without regard to its scale. In one dimension, the shape is determined by knowing the relative sizes of the various intervals between points on a line. But it is not necessary to know the length of any interval in terms of a standard scale. In order to completely determine the structure, all that is needed is to measure one interval against a standard scale, which would then allow us to relabel the coordinates in the scale of the standard unit. In this paper, we do not regard scaling as part of the calibration process but rather think of it as something that can be accomplished after the shape of the coordinate markers has been determined.

Indeterminacy. In the example of L above, self-consistency allowed us to infer the correct structure of the coordinate markers. I call this the determinate case. It is important to note that self-consistency does not always lead to determinacy but may lead to indeterminacy. Again, let this be illustrated on the line L. Suppose that, in the example above, we had begun by shifting the line, call it M this time, two intervals rightward instead of one interval. (See Figure 6 again.) In that case, the even-numbered intervals could be compared to one another, and the odd-numbered intervals could be compared to one another. But the odd- and even-numbered intervals could not be cross-compared. In this case, the procedure partitions M into two disjoint sets of intervals. Self-consistency of the *shift-by-two* procedure allows only the conclusion that all the even-numbered intervals are of one length and all the odd-numbered intervals are of one length. But we cannot determine the relationship between the two separate lengths. This is an indeterminate case. Similarly, a self-comparison procedure based upon any other single, integral shift of more than one interval (leftward or rightward)

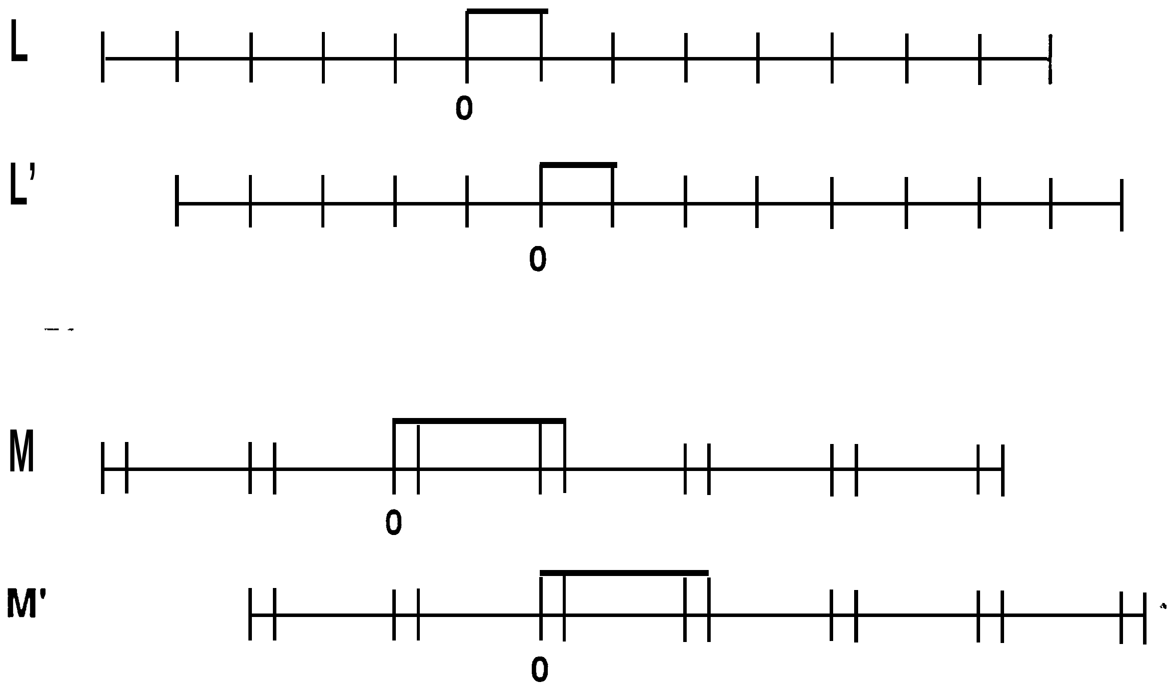

Figure 5. *Illustration of self-consistency in one dimension.* A line *L* is marked at intervals such that when a copy *L'* is shifted one interval rightward, the marks line up as shown. An obvious induction argument shows that all the intervals must be of equal length. Similarly, a line *M* is marked and a copy *M'* is shifted by the combined length of two consecutive intervals as shown. Coincidence of the marks for M and M' implies that the shorter intervals are all of equal length and that the longer intervals are all of equal length, respectively, but we are unable to compare the two lengths. In this case, self-consistency does not imply that all the intervals are of equal length.

leads to indeterminacy. A single *shift-by-n* leads to a partition of M into n distinct *equivalence classes* of intervals, in any one of which the intervals are all of one length, but it is impossible to cross-compare the lengths of intervals in any two different equivalence classes. This also is an indeterminate case if n is greater than 1.

These same ideas carry over to the circumference C. Suppose that on the circle we have set down a. finite number of coordinate markers at approximately equal intervals around the circumference. Imagine rotating the circle about its center P. We sometimes refer to this center of rotation as a *pivot point* or rotational *fixpoinf.* As in the case of the line, if we could adjust these markers in such a way as to ensure self-consistency with one rotation of the circle through an angular width of one interval, then we could inductively show that all the markers are equally spaced. Thus we would have an exactly ordered set of coordinate markers spread out at equal intervals around the circumference. This would constitute a calibration. As before, the scale could then be adjusted to a standard unit by an additional operation of comparing the length of one of the intervals to a standard scale, i.e., by measuring one interval with a standard measuring device.

Instead of comparing C with itself rotated by one angular interval, we could pick some other angle, say an angle equal to the combined width of the first two angular intervals. If the total number of intervals were even, then this procedure would allow us to implicitly compare the even-numbered intervals among themselves. Likewise, we could compare the odd-numbered intervals among themselves. But we could not cross-compare an odd-numbered interval with an even-numbered interval. Thus if we have self-consistency we could conclude only that the even-numbered intervals were all of one length, and that the odd-numbered intervals were of one length, but we could not compare the two sets of intervals. In fact, the relative lengths of the two kinds of intervals would be indeterminate, and accordingly we say that the measurement procedure itself is indeterminate. Once again we see that self-consistency may not be sufficient to calibrate the coordinate markers.

Orbits. It is important to know the kinds of procedures for which self-consistency allows us to derive a valid calibration, such procedures being called determinate, and also to know when calibration procedures are indeterminate. In the examples above we described two different kinds of procedures for self-comparison of a coordinate system. In the determinate kind we could (implicitly) compare all intervals to one another. In the indeterminate kind we wound up with disjoint subsets of intervals, in any one of which we could make implicit comparisons, but we could not make cross-comparisons between the subsets. Since mutual comparability is a crucial concept, let us introduce a term to identify a set of comparable intervals. By the *orbit* of an interval with respect to a given set of procedures,

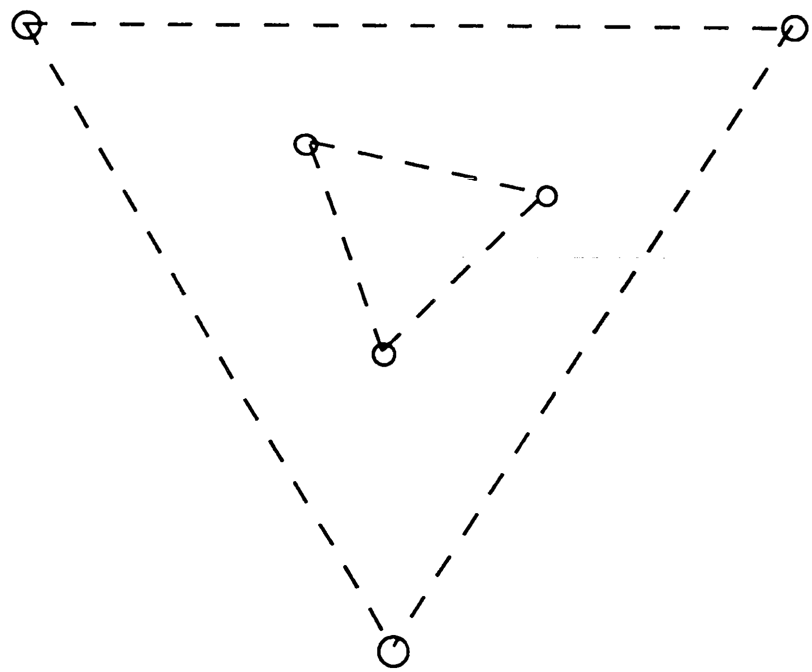

Figure 6. *A radially symmetric ball plate.* The ball plates examined in this paper are all approximately radially symmetric for some particular center of rotation and angle of rotation. Whatever angle of symmetry is chosen, it is assumed that it is possible to rotate the ball plate through that angle about one center point any number of times and to fasten it properly on the instrument's measuring plane. The balls may be of differing radii, and we assume that the radii are unknown.

I mean the set of all intervals that can be compared to that interval, directly or indirectly, in the manner illustrated above.[8] For a shift-by-n procedure on *L* there are *n* distinct orbits. It is useful to note that the situation on C is slightly more complicated. If on C there are a total of *m* intervals, and a *rotation-by-n* procedure is used, then there is exactly one orbit if and only if *m* and *n* are relatively prime, a fact that is easily proved by elementary number theory. A similar case can arise on *L* if instead of one shift, the self-comparison procedures are based upon two different shifts, say *shift-by-m* and *shift-by-n*. When self-comparison procedures yield precisely one orbit, then the procedures are said to be transitive. Alternatively, procedures that give rise to multiple orbits are said to be nontransitive. A basic principle of self-calibration can now be stated simply.

Principle of Self-Calibration: The One-Dimensional Case. Calibration can be deduced only from transitive measurement procedures; calibration cannot be deduced from nontransitive procedures.

The principle is easily seen to be correct for the simple one-dimensional examples given above. We will now proceed to show that the same principle holds in more complex two-dimensional problems.

Self-Calibration in Two Dimensions

General Considerations. Self-calibration in two dimensions will be illustrated using ball plates. The examples are used to show the limits inherent in efforts to self-calibrate by rotating a ball plate around a fixed pivot point. It will be seen that the self-calibration problem can be solved by rotating the plate around more than one pivot point. Each of the ball plates in the examples is assumed to be radially symmetric about its center, at least to first approximation. (See Figure 6.) Various degrees of rotational symmetry are considered. I do not assume that the radii of all the balls have been accurately determined by some prior measurement before attachment to the plate. I do assume, however, that the spheres are perfectly spherical so far as our Coordinate Measuring Instruments are concerned.

---

[8] In the earlier discussion, the mutually comparable intervals were referred to as an equivalence class. I prefer to use the term "orbit" because it arises naturally in the group-theoretic approach to calibration that we will take up in Part II. The orbits discussed here are comparable to orbits of points under the *action* of a rigid motion group (Armstrong 1988, 91-93; Jacobsen 1974, 69-78). I have also adopted the term "transitivity" by virtue of its similar use in the theory of group actions, yet there is an important difference. A group action on a set is said to be transitive if the orbit of an element is the entire set. When used in the group-theoretic approach to self-calibration, I do not require transitivity to be so stringent; I only require that the orbit be dense in the set. Thus my use of "transitivity" is more like "metric transitivity" as the term is used in ergodic theory (Halmos 1974, 25; Petersen 1983, 42). I will take up this issue in a sequel.

The question I raise is whether it is possible to calibrate a Coordinate Measuring Machine and simultaneously determine the configuration of the balls on a plate, merely by deductions based upon repeated measurements of the plate in various positions. In particular, is it possible to calibrate the Coordinate Measuring Machine by simply rotating the plate and remeasuring it any number of times about a fixed center? I conclude that, in general, the answer is no although there are simple exceptions! And I show why, in general, self-calibration can be achieved by rotating the plate about more than one center. The fundamental concepts that arise from the analysis here, as in the one-dimensional case, are self-consistency and transitivity. In the two-dimensional case transitivity refers to a property of calibration procedures such that each coordinate patch of a ball plate is compared, implicitly or explicitly, with every other CP of the ball plate. We shall see that transitive, self-consistent procedures enable calibration, non-transitive procedure do not enable calibration even if self-consistent.

For the sake of demonstrating principles, I will assume that at the outset we know only the approximate configuration of the ball plate, not the precise configuration. Also I assume that we can control manipulations of the ball plate well enough to rotate the plate around a perfectly fixed center.

Assumptions about Ball Plates and Coordinate Measuring Machines'. Let us suppose that we have an arbitrary ball plate and that it can be set down and measured upon the measuring plane of a Coordinate Measuring Machine. When the ball plate is first positioned on the measuring plane, we will refer to this as the ball plate's *primal* position. As explained in the section Preliminary Examples and Definitions, we can use measurements of the ball plate in its primal position to establish orthonormal frames of reference for subsequent measurements.

To be specific, suppose that the balls are labelled B_i where i runs through some index set, for example 0, $1, 2, 3,$ etc. Occasionally, B_i will also be used to refer to the center of the i^{th} ball. I assume that the following is possible. The first placement of the plate may be used to set up local orthonormal coordinate systems, one for the neighborhood of each ball. I refer to these coordinate systems as Coordinate Patch $1 (CP_1)$ for the one surrounding B_1, and Coordinate Patch 2 (CP_2) for the one covering the neighborhood of B,, and so on, respectively. (See Figure 3 again.) Henceforth, when we refer to measurements being made in CP_i we will mean measurements made by the Coordinate Measuring Machine in the region of the measuring plane that was occupied by B_i in its primal position.[9] So, even if the ball plate is moved, CP_i remains fixed on the stage, and measurements taken in CP_i are assumed to be represented in the orthonormal coordinate system that was established by the initial probing and measuring of B_i in its primal position. Later on, when we have deduced the location of the center of rotation with respect to each CP, we will again modify the local coordinate system by translating its origin to the respective center of rotation. The alleged orthonormality is an idealization. My intention is to allow some large and small motions of the plate in ways such that each ball will land within a well-defined local coordinate patch. I assume, to first approximation, that each patch is small enough that our assumptions of local linearity are reasonably correct. Since I am only discussing principles here, I also assume that measurement error and round-off are negligible.

Gross Rotations and Fine Rotations. I will deal only with ball plates that are radially symmetric to first approximation. I assume that by some means the ball plate can be rotated about a fixed center on the stage very close to the supposed center of symmetry of the plate. I then imagine two kinds of rotation of the ball plate: 1) *gross rotations* through an angle approximately equal to the angle of symmetry of the ball plate, and 2) *fine rotations, to* be defined below. In either case, I assume that each ball will fall within one of the coordinate patches, and I assume that all such rotations are taken about the same fixed rotational center. So, for example, fine rotations will move each ball along a circular arc entirely within its coordinate patch, and all such circular arcs will be centered at the fixed rotational pivot point. A gross rotation, also centered at the same fixpoint as that for fine rotations, will move each ball out of one coordinate patch and into some other coordinate patch, except perhaps for a ball that could be located in a patch that contains the center of rotation. These ideas apply generally to any number of balls (two or more) arranged symmetrically about a central point. To be specific, imagine three balls arranged at the vertices of an equilateral triangle, with each ball lying within its primal CP. A gross rotation of approximately $120°$ about a point located approximately at the center of the triangle would move each ball from its primal CP into another CI? (See Figure 7.) Once the primal CPs are determined, they remain fixed in the measurement field of the Coordinate Measuring Machine. The ball plate can then be rotated in any way, through fine angles or gross angles, with the sole restriction that each ball must settle within some coordinate patch or other.

Procedures for Locating the Origin of a Coordinate Patch. Two fine rotations (i.e., three placements of the plate arising from one original placement and two rotations about a fixed center) are sufficient to determine,

[9] We could say that the Coordinate Measuring Machine is "locally calibrated" within each CP

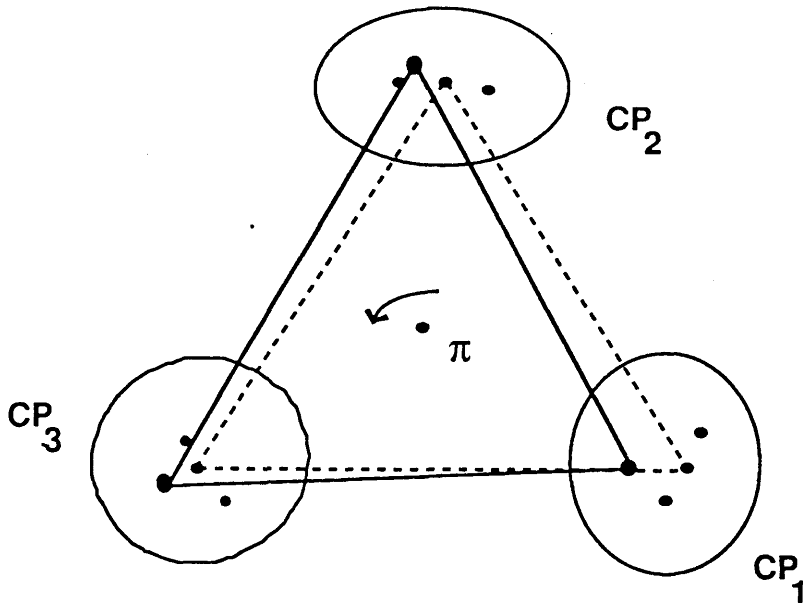

Figure 7. *Fine and* gross *angles.* A ball plate with three balls centered at the vertices of an approximately-equilateral triangle is measured in four positions obtained by rotations around a fixed pivot point π located at the approximate center of the triangle. The dashed lines represent the ball plate in its initial position, where the initial measurements were made. Two fine rotations about the pivot point were made, yielding two additional positions for measurement as indicated by the two extra small dots in each CP. Next, the plate is rotated about π through a gross angle of approximately *120°,* where it is measured a fourth time. These procedures are not optimal for self-calibration, but they make it easy to compute the center of rotation π in the local coordinates of each coordinate patch, to compare the scales of all three coordinate patches, to determine the exact shape of the dashed triangle, and, finally, to conform the three coordinate patches into one consistent Cartesian system -- in other words, to calibrate the measuring machine on the union of the three coordinate patches, as explained in the text.

in terms of each local coordinate system, the center and radius of the circular arc traced by each ball within its respective coordinate patch. This is a simple consequence of the fact that three points on the circumference of a circle determine the circle. Thus by taking measurements of a ball in three placements (i.e., primal position plus two fine rotations) within a coordinate patch, we can determine the center of rotation, referenced to the local CP coordinate system (although this computed center may lie far outside the "range" of the local CP), and we can also compute the radius and respective angles of rotation within the respective CI?[10] Since we can do this for each of the balls, we can determine the center and radius of rotation for each ball in the respective local coordinate system. Having done so, we translate each local coordinate system so that its origin coincides with the local rotational fixpoint as computed within that local coordinate system. Henceforth, we shall assume that the local coordinate system in each CP is orthonormal and centered at the center of rotation.

Procedures for Determining a Common Scale. To sum up the situation, we now have a radially symmetric array of distinct coordinate patches (one corresponding to each ball, three in the example). Each is coordinated by an orthonormal coordinate system, and each such system has its origin lying at the unique center of rotation of the ball plate. But the relative scales and orientations of these coordinate systems is, as yet, unknown. In order to calibrate the Coordinate Measuring Machine, or rather to calibrate the subset of the Coordinate Measuring Machine covered by the given set of coordinate patches, we must relate each of these local coordinate systems to one another. We shall achieve this goal in several steps. The first step is to make a gross rotation, moving each ball into a neighbor coordinate patch. In the new coordinate patch we can remeasure each ball. A comparison of the radius of the ball B_1 computed in the first coordinate patch and the radius computed in the second coordinate patch permits us to rescale the second coordinate patch to match the scale of the first.[11] Note that if B_1 has moved from CP_1 to CP_2 , and B_2 has moved from CP_2 to CP_3 , then we

---

[10] Because the ball plate is rigid, the respective rotation angles as computed in the various CPs must be equal; the corresponding radii of rotation, however, may vary from CP to CP since the $\Delta B_1 B_2 B_3$ is not necessarily exactly equilateral, the rotation point is not necessarily exactly at the center of the triangle, and each CP may have a different scale relative to the other CPs. As an idealization, it has been assumed that each local CP coordinate system is perfectly orthonormal and that there is no measurement error or roundoff; however, in a practical application of the ideas of this analysis one would have to take into account departures from the ideal. For example, the center of rotation as estimated from the measurements within each CP should be expected to vary quite sensitively as a function of the errors.

[11] Instead of comparing the radius of a ball as measured in each of two CPs, we could just as easily compare the distance of the center of the ball from the center of rotation of the ball plate as measured in each of the two CPs.

can also indirectly compare the scales of CP_1 and CP_3, and so on. So it is quite possible to achieve a cascade effect, in which one gross rotation permits inductive comparison and equilibration of the scales of all the coordinate patches within the entire cascade. For example, if the ball plate consists of n balls arranged at the vertices of a regular polygon, then one gross rotation approximately equal to the symmetry angle of the polygon can allow a comparison and equilibration of all the coordinate patches. This procedure is transitive in the sense defined above -- the procedure of rotating the plate once and remeasuring allows implicit comparison of the scales of all the CPs by application of the principle of logical transitivity.. Note, however, that if one additional ball were positioned outside the regular polygon of the first n coordinate patches, for example at the center of the polygon, the procedure would not be transitive, because no comparison could be made between the odd-ball and the polygonal ones. This will be explained more fully below. For now, assume that the scale of each of the three CPs is adjusted to equal the scale of CP_1. Thus, henceforth we may assume that all three CPs are orthonormal and share a common origin and scale. Next, we would like to determine the relative orientation angles of the three CPs.

<u>Examples of Orbits.</u> At this point, I want to introduce another useful term. Above I referred to the notion of a cascade. More precisely, consider a particular ball, say B_i in CP_i and observe where it moves under a particular gross rotation, say of angle θ. Suppose that B_i moves into the coordinate patch CP_j previously occupied by B_j, which in turn moves to, say, CP_k previously occupied by B_k, and so on. It is this sequence of coordinate patches that I meant to imply by the term "cascade." Notice that, for a ball plate with coordinate patches all lying at the vertices of a regular polygon, a cascade must eventually turn back upon itself, because the number of coordinate patches is finite. A cascade is a closed, linked list of coordinate patches, which we refer to descriptively as the *orbit* of B_i *generated by the gross rotation through angle* θ. Also notice that if a plate undergoes a gross rotation, then each coordinate patch will belong to one and only one orbit; thus the orbits generated by that rotation partition the coordinate patches into distinct, disjoint orbits. As already mentioned, an orbit generated by a single rotation consists of a sequence of coordinate patches that lie approximately at the corners of a regular polygon centered at the center of rotation. But other procedures can give rise to different kinds of orbits. For example, suppose that the ball plate consists of four balls arranged at the vertices of a square, and that the procedure is to measure the plate, then rotate the plate through 180° and remeasure. This procedure yields two orbits -- each consisting of two diametrically opposed CPs. For another example, suppose the ball plate consists of eight balls, four disposed on the vertices of a small square, and four at the vertices of a larger concentric square. Now, if the procedure were to

rotate the plate once through 180 °, there would be four orbits: two at the inner square and two at the outer. (See Figure 8.) If, instead, the procedure were to rotate the plate once through *90°*, there would be two orbits. If we added a central ball to this configuration, the first procedure would yield five orbits and the second would yield three orbits.

It is obvious now that we can compare scales among coordinate patches lying within an orbit. It is also true, but perhaps not so easy to see, that inter-orbital comparisons are impossible without additional procedures. Thus transitivity exists only within an orbit, not between distinct orbits. This fact will be illustrated in the examples below. It is important enough to give it a name.

Principle of Self-Calibration: The Two-Dimensional Case. Procedures for self-calibration can relate only coordinate patches that fall within an orbit. The distance relationships and angular orientations between distinct orbits are indeterminate (for an interesting variation, see Exercise 4 below)? Full calibration can only be achieved with transitive procedures. In particular, it is impossible to calibrate a general ball plate without rotating the plate around more than one pivot point, since a procedure of rotations about a single pivot point cannot be transitive, except for the very special case in which all balls lie on the vertices of a regular polygon. In general, transitivity (and hence calibration) can only be obtained by rotations about more than one pivot point.[13] This principle is demonstrated in the examples that follow.

The Three-Ball Plate. We will progress through this section by stages. Our objective is to determine the precise configuration of a three-ball plate by

---

[12] In two dimensions, we say that the configuration of a ball plate is "determinate" if the internal angles (see glossary) between all the ball centers are well-defined by the measurement procedures. Note that these angular relationships determine the shape of the configuration, but they do not determine the scale. If the measurement procedures do not determine all internal angles unambiguously, then the ball plate is said to be indeterminate. By extension, the measurement procedures themselves may be said to be determinate or indeterminate, depending upon whether they do or do not determine the shape of the ball plate. Finally, the coordinate system of the Coordinate Measuring Machine is said to be determinate or indeterminate, depending upon whether the procedures used to measure the ball plate are likewise determinate or indeterminate. To state it succinctly, procedures for measuring a ball plate on a Coordinate Measuring Machine are indeterminate if two differently configured ball plates (i.e., ball plates with distinctly different internal angles) measured on two different CMMs with two different coordinate systems can yield identical sets of measurements; procedures are determinate if they are not indeterminate.

[13] So far we have only allowed rotations of a ball plate around a single fixpoint. In this case an orbit is comparable to a regular polygon centered at the fixpoint. However, if more than one rotational center is allowed, orbits may be more complicated collections of coordinate patches. Because transitivity is a desired feature of self-calibration procedures, we should in fact want there to be exactly one orbit embracing all the coordinate patches.

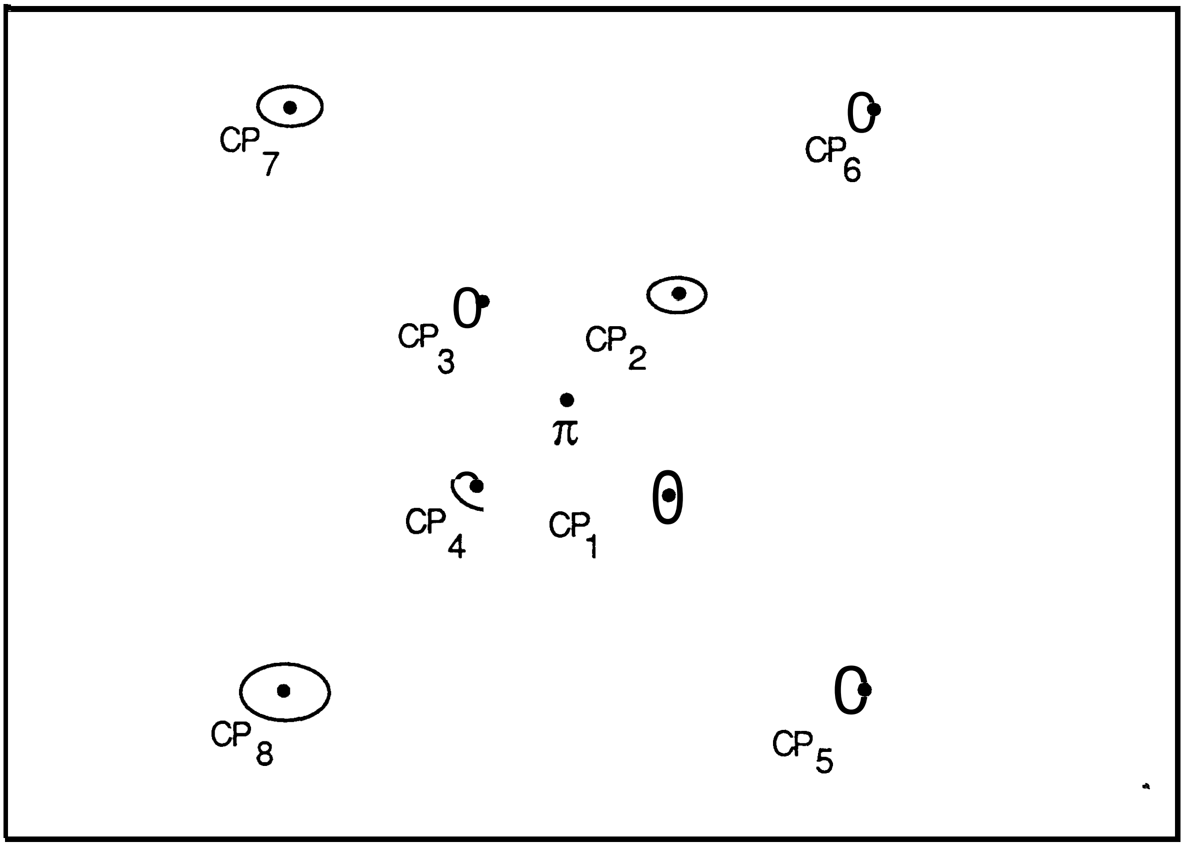

Figure 8. *Procedures* can *yield multiple orbits.* The unlabeled centers of eight balls are shown, one within each of eight coordinate patches. The bails (and CPs) are arranged in a pattern of $90°$-rotational symmetry. A fixed rotation point π near the center of the pattern is shown. Each ball is measured in its given position, then the plate is rotated counterclockwise about π through approximately $180°$ in such a way that each ball lands within an existing CP, where it is remeasured. This procedure results in four distinct orbits: $CP_{1\,3}$, $CP_{2\text{-}4}$, CP_5, , and $CP_{6\,8}$, where $CP_{i\text{-}j}$ denotes CP_i and CP_j.

taking measurements as the plate is rotated about a single pivot point and to determine a satisfactory Cartesian coordinate system for the union of the three associated coordinate patches. To achieve the latter we must deduce the geometric relationships among the three coordinate patches. We will, in effect, set each of the coordinate systems of the three CPs, respectively, into a single Cartesian framework. I refer to this process as conforming the three coordinate systems. The resulting Cartesian coordinate system will be referred to as the conformed coordinate system. Note that, in the previous discussion, we have made a start in this process by translating the origin of each CP to the center of rotation of the ball plate and rescaling each of the CPs to match the scale of the first CP. Here we are using an orbit of three CPs as an example, but the ideas apply as well to all the CPs in an orbit of any size.

Special Assumptions. I assume that the balls lie centered at the vertices of a triangle that is nearly equilateral and that the pivot point lies near but not necessarily at the center of the triangle. (See Figure 7 again.) I assume also that the stage of the Coordinate Measuring Machine permits 120 O-rotations of the plate. These restrictions are made to ensure that, as the triangle is rotated about the fixed pivot point, the three vertices will traverse arcs of a circle, or of three concentric circles of nearly equal radii, and that the 120 O-rotations will result in near-overlays of the triangle upon itself. Two sets of rotations and measurements will be employed. First, the three balls are probed and measured in the primal position of the plate in order to establish the orthonormal coordinate system for each respective CI? Second, two fine rotations are made about the fixed rotational center; after each such rotation, the balls are re-measured to determine coordinates for the center of rotation referenced within the coordinate system of each CP, respectively. The rotations are made small enough to ensure that each ball remains within its respective CI? At this point, I have made nine measurements for the centers of the three balls -- three center positions for each ball. Finally, one gross rotation of approximately 120 ° is made, and each ball's new position and radius is measured in its respective new local CP.[14]

The Deductions thus Far. Let's take stock of what we know from the initial measurements of the balls, i.e., the measurements taken before we make the 120 ° rotation. We have assumed that each CP has its own

---

[14] The first three positions of the ball in its primal coordinate patch were sufficient for determining the center of rotation and the distance from the center of the ball to the center of rotation within the coordinate system of that CP. Thus, when the next move is made, namely, the gross rotation that swings each ball out of its primal neighborhood into an adjacent coordinate patch, we can compute not only the distance of the ball from the center of rotation in the local coordinate system, thereby obtaining a radius that can be compared with the radius calculated in the primal position, but also an angle with respect to the center point of rotation between the current position of the ball occupying the coordinate patch and the first position of the previous ball in the same coordinate patch. (See Figure 9.)

orthonormal coordinate system. The three measured centers of the local ball are then sufficient to determine, within each respective coordinate system, the center of rotation, the two fine angles of rotation, and the radius of the circular arc on which the ball has rotated. These factors are thus known within the local orthonormal coordinate system of the local CP. The same is true for each ball in its primal coordinate patch, respectively. As already mentioned, we translate the origin of each local coordinate system to the respective center of rotation within that system. Until we are able to compare the three scales, we can't know the three radii relative to each other, but in any case we do know that the center referenced in each of the coordinate systems is the one identical center of rotation of the ball plate. Moreover, because the plate is rigid, we know that the corresponding fine angles of rotation are equal from one patch to the next. Thus far, each ball has remained within its own coordinate patch.

Now let's see what we gain from the gross rotation of 120 °. Note first that each ball has moved from its original coordinate patch into a different patch. By measuring each ball's radius in the new coordinate system, we are now able to compare the scale of the new CP to the scale of the original CP. Since the procedure yields a single orbit containing all three CPs, we are therefore able to compare (and equalize) the scales of all three CPs. We make this comparison and further modify each of the three local coordinate systems so that they all conform to the same scale.[15] Thus we now know the precise relative distance of each ball from the sole center of rotation. Let us express these radii in a common scale as $r_1, r_2,$ and $r_3,$ where

r_i = radius from center of plate to center of B_i, for $i = I, 2, 3$.

<u>Equations for Determining Shape.</u> To determine the precise shape of the triangle formed by the centers of the three balls, consider the fact that a straightforward application of analytic geometry can be used to compute the angular offset that each ball (after the gross rotation) makes from the original position of the ball that previously occupied the same Cl?, measured as a counterclockwise angle of rotation about the fixed center of rotation. (See Figure 9.) We refer to these angular offsets, measured in the counterclockwise sense, as:

1 5Henceforth, we assume that the three CPs have coordinate systems that are 1) orthonormal, 2) originate at the center of rotation of the ball plate, and 3) have uniform scale. What we have not determined yet are the relative orientations of the three coordinate systems, but that will follow directly once we determine the shape of the triangle formed by the centers of the three balls, respectively.

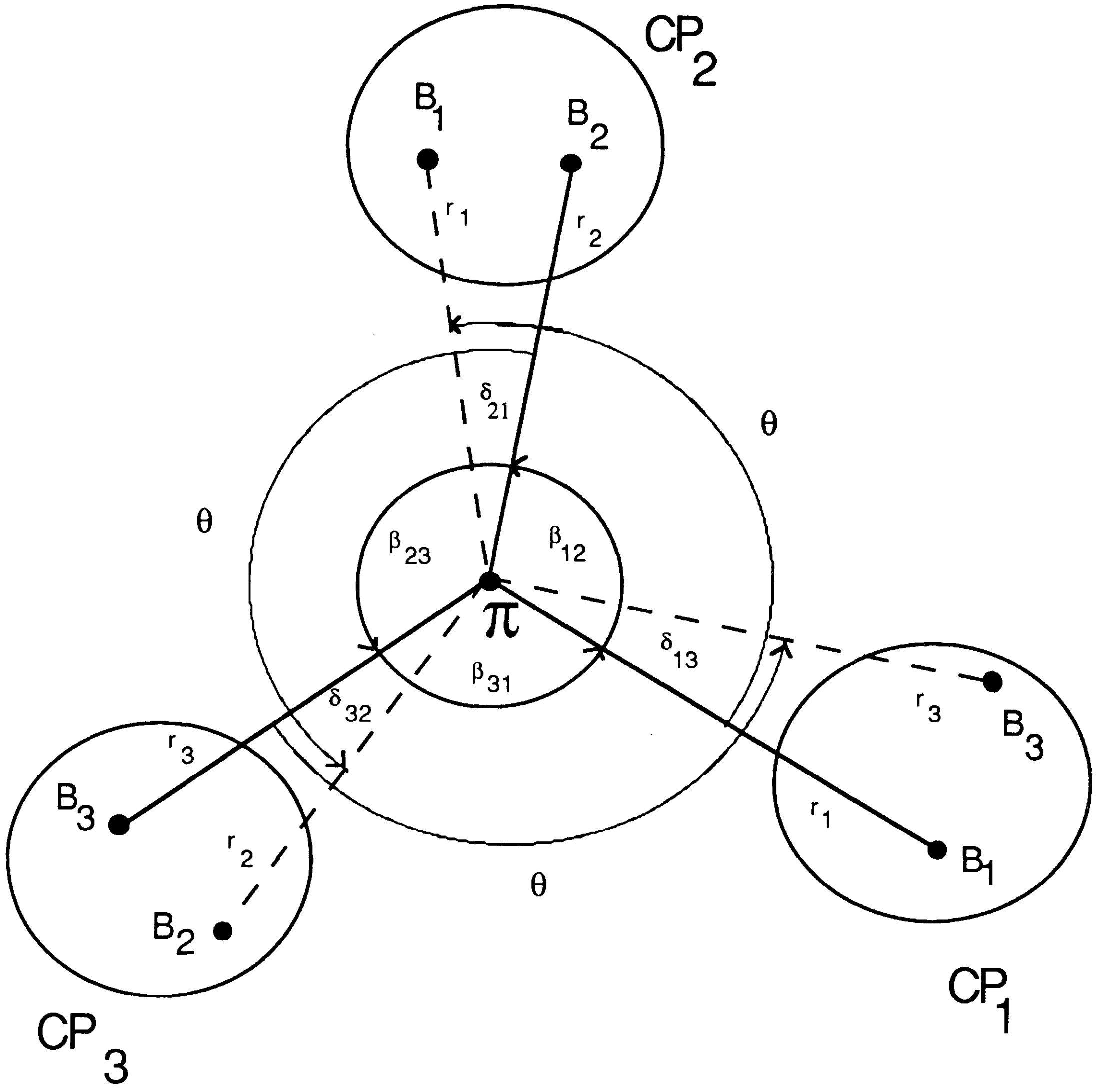

Figure *9. Variables used to determine the shape of a ball plate.* Three CPs are shown, centered approximately at the vertices of an equilateral triangle. A rotation point π lies near the center of the triangle. A ball plate with three ball centers is shown in each of two positions. Each center B_1, B_2, and B_3 is shown in its first position (with radius from the center of rotation π illustrated as an unbroken line) and in its second position after a counterclockwise rotation of angle $\theta \cong 120°$ (with radius from π illustrated as a broken line). For each pair of balls, say B_i in its first position and B_j in its second position, shown within a given CP, the angle of offset measured in a counterclockwise sense from B_i in its first position to B_j in its second position is δ_{ij}. Note that the shape of the triangle with vertices B_1, B_2, and B_3 is determined by the center π, the radii r_1, r_2, and r_3, and the internal angles β_{12}, β_{23}, and β_{31}.

δ_{13} = measured angle of offset from old B_1 to new B_3 in CP_1 ,

δ_{21} = measured angle of offset from old B_2 to new B_1 in CP_2 ,

δ_{32} = measured angle of offset from old B_3 to new B_2 in CP_3 .

Recall now that the ball plate is a rigid object, and let

β_{12} = angle from center, from B_1 to B_2, ,

β_{23} = angle from center, from B_2 to B_3 ,

β_{31} = angle from center, from B_3 to B_1 ,

θ = angle of gross rotation of plate (approximately $120°$).

Because the ball plate is a rigid object, we deduce the four equations:

$$\beta_{12} + \beta_{23} + \beta_{31} = 360°,$$
$$\beta_{12} + \delta_{21} = \theta,$$
$$\beta_{23} + \delta_{32} = \theta,$$
$$\beta_{31} + \delta_{13} = \theta.$$

These equations are easily solved for θ , β_{12} , β_{23} , and β_{31} in terms of δ_{13} , δ_{21} , and δ_{32}. This information, combined with knowledge of the radii r_1, r_2, and r_3, permits us to determine the precise shape of the triangle $\Delta B_1 B_2 B_3$ and the angle θ through which the triangle was rotated from the first to the last position. I leave this as an exercise for the reader. Moreover, we can now easily place the three local coordinate systems into a common framework. For example, we could transform coordinates for CP_2 and CP_3 to conform with the coordinates of CP_1. Let us refer to this common coordinate system as the $CP_1/CP_2/CP_3$ conformed coordinate system, or simply as $CP_{1\text{-}2\text{-}3}$.[16]

---

[16] To obtain a conformed coordinate system, let ϕ_i refer to the counterclockwise angle made by the center of B_i from the X-axis in its original position in CP_i , for i = 1,2,3, respectively. Let ϕ_i' be the counterclockwise angle that B_i <u>should</u> make with the X-axis, if CP_i were conformed to CP_1 , for i = 2,3. respectively. Note that

$$\phi_2' = \phi_1 + \beta_{12}$$

Note that, by the act of conforming the three coordinate systems, we have simultaneously calibrated both the ball plate and the Coordinate Measuring Machine, at least within the union of the corresponding coordinate patches.

Role of Self-Consistencv in Deriving Orbital Coordinate Svstem. We were able to link the three coordinate patches together because we were able to compare the three scales and use the rigidity of the ball plate to deduce the shape of the triangle, and this was because the three coordinate patches fell within a single orbit.[17] It is instructive to consider how self-consistency played a role in the derivation. By the hypothesis that we are operating in a euclidean universe, we assert that there is essentially only one way to coordinate a single artifact with Cartesian coordinates, at least so far as measuring its true shape is concerned. Although the particular Cartesian reference frame may be chosen any number of ways, the difference in image between any two such frames must simply be a matter of uniform scale, a rotation, and a translation, but the shape of the object must be identical for all such frames, no matter where we place the object to be measured. In our derivation of a conformed Cartesian reference system, we used the methods of analytic geometry to proceed constructively from the assumption that we were dealing with a rigid object and showed that the procedures were sufficient to determine the shape of the object from the given measurements. Thus we may say succinctly that we used self-consistency and the euclidean axiom to derive the shape of the ball plate directly from the measurements.

A Four-Ball Plate, Used to Illustrate Indeterminacy. At this point we will examine the effect of introducing an additional ball at the rotational center of

---

$$\phi_3' = \phi_1 + \beta_{12} + \beta_{23} .$$

It follows that conformation of CP_2 to CP_1 requires a counterclockwise rotation of CP_2 coordinates of magnitude $\phi_2' - \phi_2$, and conformation of CP_3 to CP_1 requires a counterclockwise rotation of CP_3 coordinate of $\phi_3' - \phi_3$. These counterclockwise rotations of coordinates are equivalent to equal clockwise rotations of coordinate axes for each respective CP.

[17] We have worked out the shape of a regular polygon formed by three balls, but the same reasoning would allow us to work out the shape of a regular polygon formed by any number of balls. In such a case, we would have had to make a gross rotation equal to the symmetry angle of the polygon (120° in the case of a three-ball plate). And we would have deduced equations similar to those above but there would have been $n+1$ of them, with $n+1$ unknowns, where n is the number of vertices of the polygon. An interesting exercise that the reader may use to test understanding of the ideas thus far, would be to work out the case of a two-ball plate ($n = 2$) as the simplest example. Don't assume that the center of rotation lies precisely on the line between the two points.

the three-ball plate. With the additional ball, the same procedures as before will give rise to two orbits rather than just one. We will show that by virtue of the additional orbit, the problem becomes doubly indeterminate: We can not determine the relative scales of measurement between the two orbits, nor can we determine the relative orientation between the two orbits. The dilemma will be seen to be a very general one that applies to comparison of any two distinct orbits.

To obtain the four-ball plate we add a central ball and coordinate patch B_0 in CP_0, respectively. It seems intuitive that we cannot compare the scale of the central coordinate patch with the scales of the three outer CPs, nor could we have determined the relative orientation of CP_0 and the conformed $CP_{1\text{-}2\text{-}3}$ coordinate system, derived as in the three-ball plate in the preceding section. Although this may seem self-evident, it requires proof. Before giving the proof, let's be explicit in what it means to say that the four-ball plate consisting of B_i, for $i = 0,1,2,3$ is of indeterminate shape. I mean that it is possible for two ball plates of distinctly different shapes to be measured by two different measuring devices and yet to obtain identical measurements for the two ball plates.

Shape Defined more Carefullv. First, let's be more specific about what is meant by the shape of an object. In plane geometry two triangles are said to be *similar* if their corresponding angles are equal. If the two triangles are also of equal size they are said to be congruent. I say that two triangles are of the same shape if they are similar but are not necessarily congruent. The classical test for congruence of two triangles is to move the one through space to see whether it can be superposed exactly on the other. Another way of testing for congruence is to measure the two triangles in a Cartesian coordinate system and see whether the coordinates of the vertices of the one triangle can be transformed into the coordinates of the vertices of the other by a rotation and translation of coordinates. Similarity of two triangles requires that there be a rotation, translation, and uniform scaling that carries the one set of vertices into the other. The same ideas carries over to more general geometric figures. Thus we can test two geometric objects for congruence by testing whether the one can be moved rigidly into perfect coincidence with the other, or we can measure the objects in a Cartesian coordinate system and see whether there is a rotation and translation that transforms the coordinates for the one object into the coordinates of the other. To test for similarity of the two objects, we allow not only rigid motions (rotation and translation) but also uniform scaling of objects. For two-dimensional patterns of points, there is a convenient way for us to determine whether two different patterns of points have the same shape: simply compare the corresponding *infernal angles.* The internal angles of a pattern of points is the set of all the angles of the triangles formed by any three points of the pattern. If the points of two two-dimensional patterns of points can be put into correspondence with one another in such a way that the corresponding internal angles are equal, then the two patterns are similar. If such a correspondence is impossible, then the patterns are not similar.

What I want to do now is show that if the measurement procedures for a four-ball plate are as above, then we cannot determine the shape of the ball plate. Moreover, the same is true even if additional measurements are made of the ball plate by rotating the ball plate around the one fixed center of rotation and remeasuring it. There are different ways to show this, but perhaps the simplest and most direct way is to demonstrate that we can change the shape of the given ball plate and the coordinate system of the given Coordinate Measuring Machine in such a way that, for any rotation of the one ball plate about its center and the same rotation of the other about its center, the new ball plate will yield the same measurements on the new measuring machine as the given one does on the given machine. What this means in practice is that, given one such ball plate and Coordinate Measuring Machine, there is no way to determine from just the measurements which ball plate we are examining.

Constructive Demonstration of Indeterminacy. To begin, assume that we are given a four-ball plate and that it has been measured on a given Coordinate Measuring Machine by the series of fine rotations and gross rotations described above for the three-ball plate, and that we have derived a conformed coordinate system, call it $CP_{1\text{-}2\text{-}3}$, for the outer orbit. Assume also that we have gone through the steps described previously to render the inner orbit CP_0 into Cartesian coordinates, that we have computed the center of rotation in terms of CP_0, and that we have translated the origin of CP_0 to that center. Finally, assume that the ball plate is fastened to the Coordinate Measuring Machine in an initial position, with each ball resting in one of the four coordinate patches. Now, we can modify any portion of the ball plate that we choose. Let us modify just the portion in the outer orbit by anchoring it at the center of rotation, scaling it up (or down) uniformly by any factor $s > 0$ and rotating it counterclockwise by any angle ϕ. We do all of this while keeping the inner orbit fixed. At the same time that we modify the outer orbit of the ball plate, let us also correspondingly modify the coordinate system of $CP_{1\text{-}2\text{-}3}$, by increasing the scale by the factor s and rotating the axes counterclockwise by the angle ϕ. We leave the coordinate system of CP_0 unchanged. The net effect of these modifications of the given ball plate and Coordinate Measuring Machine is that the new measuring machine will assign the same coordinates to the new ball plate that the original machine assigned to the original ball plate. Moreover, if the original ball plate is measured again after rotating it through any angle about the center of rotation, such that each ball remains in one of the four coordinate patches, then the same measurements will be obtained by the new machine after the new ball plate has been rotated through the same angle. (See Figure 10.)

Note that our modifications of the ball plate by stretching and rotating one orbit with respect to another has allowed us to change the shape of the ball plate. The corresponding modifications of the Coordinate Measuring Machine have produced a machine that will give the same measurements as the original machine for comparable rotations of the original ball plate. Without additional information, we cannot simply observe the measurements of one of these two machines and know which one it is. It is in this sense that I say that the shape of the ball plate is indeterminate. By extension I say also that the measurement procedures themselves are indeterminate. It is clear that this same construction can be applied to other radially symmetric ball plates and measurement procedures that entail multiple orbits and rotations about only one fixed center.

A Non-Constructive Demonstration Based on Self-Consistencv. The demonstration of indeterminacy given above is constructive. It shows how to actually make a ball plate of different shape than the original that yields identical measurements. There is another approach that reveals indeterminacy more directly in the light of self-consistency. We ask, given a radially symmetric ball plate, a measuring machine, and a set of measurement procedures entailing multiple orbits and rotations about only one center, can there be more than one shape of ball plate that is consistent with the measurements? In other words, does self-consistency allow us to determine the shape? It is sufficient to consider just two orbits. Suppose that by some means each of the two orbits has been conformed into a Cartesian coordinate system, respectively. The invariance of the euclidean metric under rigid motions implies that the images in machine coordinate space of the balls in each orbit will be consistent for any rotation that keeps the balls in the appropriate orbit. But what about the images of subsets of the balls that span the two orbits, say for example, one ball center P in the outer orbit and one ball center Q in the inner orbit? Is there a unique scaling factor s and unique angle of rotation ϕ for the inner orbit to conform it to the outer orbit, such that all inter-orbital images are self-consistent? The negative answer follows directly from the invariance of the euclidean metric. Let $E(P,Q)$ be the distance between points P and Q. where the points are represented in Cartesian coordinates. Let the expression $sQ(\phi)$ denote multiplication of the coordinates of Q by s followed by counterclockwise rotation of the coordinates through the angle ϕ about the fixed center of the ball plate. For P and Q fixed in the ball plate, and fixed factors s and ϕ, $E(P, sQ(\phi))$ is invariant under rigid motions, in particular under rotations about the rotational center of the ball plate. But the argument $sQ(\phi)$ can be interpreted two ways. It can either represent a point different from Q but obtained from Q by the scaling and rotation of coordinates described above, or it can represent the coordinates for

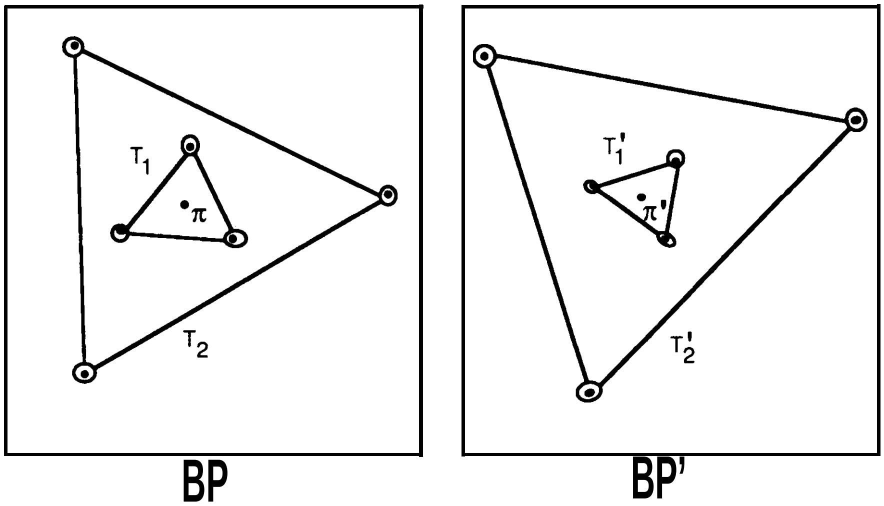

Figure 10. *Illustration of indeterminacy.* An (approximately) radially symmetric ball plate BP, consisting of six balls arranged in two concentric triangles, is mounted on the measurement plane of a Coordinate Measuring Machine. The centers of the balls are shown surrounded by their respective coordinate patches, unlabeled. BP is measured in that position, rotated, and measured again, repeatedly, using a strategy of fine rotations and gross rotations as described in the text. All rotations are about the pivot point π located approximately at the center of the ball plate and are such that each ball falls within one of the illustrated coordinate patches. As explained in the text, procedures of this kind permit determination of the precise shapes of the two triangles, but if the procedures involve only rotations about one point π, then it is impossible to determine the relative scales and orientation angle between the two triangles. As illustrated by BP', a different ball plate composed of two triangles similar to the ones in BP but scaled and oriented differently about π', and measured by a Coordinate Measuring Machine with appropriately rescaled and re-oriented coordinate patches, could yield exactly the same measurements as those obtained from the procedures for measuring BP. In fact, any arbitrary rescaling and re-orientation of the two triangles about π could be made to yield the same set of measurements for the same set of procedures by employing a correspondingly altered Coordinate Measuring Machine. Thus it is impossible to derive and verify the correct shape of BP from such measurements alone. For this reason, the prescribed measurement procedures are said to be indeterminate. The indeterminacy arises because the prescribed procedures are non-transitive, i.e., they give rise to more than one orbit of coordinate patches, namely the inner three CPs and the outer three CPs, respectively, in this example.

Q in a Cartesian system that is obtained from the Cartesian system in which Q is represented, by scaling the coordinates of Q by s and rotating Q by ϕ. The first interpretation implies that the expression for $E(P, sQ(\phi))$ is invariant under rotations of the ball plate, so it is invariant according to either interpretation. Under the latter interpretation, the invariance of $E(P, sQ(\phi))$ under rotations of the ball plate shows that self-consistency holds between the inner and outer orbits for any arbitrary scaling and rotation of the inner orbit relative to the outer. Thus, the measurement procedures do not permit **a** unique determination of s and ϕ.

Before turning to the next example, let us observe again where self-consistency was used in calibrating the three-ball plate. We used self-consistency implicitly in the assumption that the three-ball plate is a rigid object. The presumed invariance of shape under rotations was used to set up the equations for the rs, βs, and θ.

Multiple Orbits Illustrated by A Seven-Ball Plate. To the three-ball plate considered above, add one ball near the center of the equilateral triangle. Also, at a radius much greater than r_1, r_2, and r_3, add three more balls on the vertices of another equilateral triangle. Let the outer equilateral triangle be centered near the center of the inner triangle, i.e., close to the center of the innermost, central ball. Call the inner ball B_0 , and the outer three balls, B_4 , B_- , and B_6 , respectively. Now apply the same fine and gross rotations that were applied in the previous example. Note that the orbit of CP, is itself, that CP, , CP, , and CP_3 comprise an orbit, as before, and that CP_4 , CP, , and CP, constitute a distinct new outer orbit. Exactly as before, the configuration of B_1 , B_2 , and B_3 can be accurately determined in a common coordinate frame conforming CP, , CP_2 , and CP, . Similarly, B_4 , B_5 , and B_6 can be determined within a common coordinate frame conforming CP, , CP_5 , and CP,. But the relative scales of these two outer orbits is indeterminate. And so is the relative relationship of scales compared to CP,, since the three orbits **are** pairwise disjoint. Not only the scales of the three orbits but also the relative orientation angle of each orbit with respect to the others is indeterminate, as argued in the previous example. In particular, the internal angles of the pattern of ball centers are ill-defined. We conclude that these rotations about a single pivot point are inadequate to determine the configuration of the seven-ball plate. Measurement procedures that involve rotation of a ball plate around one pivot point are indeterminate, in general.

It is interesting now to reflect on the situation. Within each orbit, the shape of the pattern of ball centers is well defined. But the scale and angular relationships between the patterns in distinct orbits are indeterminate. Hence the shape of the overall pattern of ball centers is indeterminate. The orbits are concentric about the rotational fixpoint, but their relative distances from the fixpoint are ill-defined. The orbits may be thought of as *wheels-within-wheels,* and the relative orientation of each wheel to the one surrounding it is ill-defined.

Rotations about More than One Pivot Point. An orbit generated by a single rotation includes only coordinate patches that lie approximately on the vertices of a regular polygon. Similarly, an orbit around some other point would be another "polygon," with a different center. If two such polygons happen to overlap at a coordinate patch, then the two orbits would merge into one. Thus, when measurement procedures involve rotations about more than one pivot point, the resulting orbits need not correspond to regular polygons. Such an expanded orbit, generated by two rotations, can be considerably more extensive than a single-centered orbit and may in fact include all the coordinate patches of the ball plate. In that case, the techniques demonstrated in the foregoing examples may be extended to conform all the coordinate patches to obtain an accurate determination of the shape of the entire ball plate and a calibration of the measuring instrument, at least on the union of the coordinate patches. An exhaustive analysis of all the possibilities would carry us beyond the scope of this paper, but see the exercises in the next section. We will return to the subject in a later paper.

Summary, Conclusions and Exercises

Summary. We have made a discursive tour of some basic concepts of self-calibration. The purpose has been to suggest vocabulary and a framework for discussing problems in two dimensions, without going into the group-theoretic arguments of Raugh (1985), and to illuminate the importance of self-consistency and transitivity in self-calibration procedures. Emphasis was placed on exposing the basic principles of self-calibration, not upon optimization of procedures to calibrate a measuring instrument. Some one-dimensional problems were discussed to illustrate determinate and indeterminate measurement procedures and to introduce self-consistency and the complementary concepts of orbit and transitivity in the simplest settings. Next we turned to the more subtle problem of self-calibration in two dimensions. To stay on ground that is familiar to students of metrology, the two-dimensional problems were formulated in terms of ball plates and Coordinate Measuring Machines. I took the liberty of idealizing problems by assuming, among other things, that a Coordinate Measuring Machine could be adjusted to yield orthonormal coordinates in local coordinate patches and that there was no measurement error. I also assumed that the balls on a ball plate were perfectly spherical and that a ball plate could be rotated more than once about one perfectly fixed point. I did not assume that all the balls of a ball plate were of identical diameter or that the relative diameters were known beforehand.

Conclusions. The major conclusion is that self-consistency and transitivity are necessary conditions for self-calibration. In particular, each coordinate patch must be compared either directly or indirectly with every other coordinate patch in order to determine relative scales, relative orientation angles, and a common positioning of the origin. An important result is that the shape of the pattern of balls that fall within a given orbit can be determined by appropriate measurement procedures, but the relative scales and orientation of distinct orbits is indeterminate; thus, a ball plate that is measured only by rotations about a single point can be thought of as a pattern of "wheels within wheels," so that each wheel (i.e., orbital pattern) is well defined, but the relative radii and orientation angles between the "wheels" are unknowable. (A variation of this theme is given in Exercise 4 below.)

A Brief Comparison with Raugh (1985). The approach taken in this paper is very different from that of the earlier paper (Raugh, 1985). That paper will not be reviewed here. But I want to sketch some connections between the methods of this paper and the methods of the earlier one. The earlier paper introduced the concept of a *fixpoint laffice.* By appealing to group-theoretic arguments, it was shown that self-consistent measurement procedures give rise to a lattice of points that can be calibrated correctly by those measurement procedures. It turns out that there is only one fixpoint for procedures that involve rotation about a single pivot point, hence it was shown that self-calibration is impossible for that case. It was also shown that measurements of a grid plate in three positions that involve rotations about two distinct pivot points can yield a correct calibration. The reason given in the earlier paper relates to the fact that the fixpoint lattice is dense throughout the measurement space. In a sense that will be explained in a later paper, the fixpoint lattice corresponds to the orbits of the present paper, and a dense lattice corresponds to transitive measurement procedures. In the current paper, the procedures for self-calibration of each separate orbit involved measuring the ball plate first in each of four distinct placements (involving two fine rotations and one gross rotation around a single pivot center). A simple exercise is suggested below to show that transitivity can be effected by rotation about a second center and that full calibration can be attained by repetition about the new center of the same procedures that were used with respect to the original center of rotation. Accordingly, if the first four placements were followed by additional similar rotations about a second center, then a total of eight distinct placements of the ball plate would be required for calibration. These procedures are straightforward, but they are certainly not optimal. An interesting practical problem would be to determine the smallest number of placements necessary to calibrate various kinds of ball plates.

The earlier paper used self-consistency in a direct, coordinate-free way to derive necessary and sufficient conditions for self-calibration. The use of Cartesian coordinates was brought in only at the very end, to solve the practical problem, not to develop the theory. The conclusions of the current paper are weaker, but they have the advantage that they were derived using constructive methods and Cartesian coordinate systems in ways that are, for the most part, quite straightforward. Sufficient conditions for self-calibration of radially symmetric ball plates were demonstrated. Both transitivity and self-consistency were shown to lie at the heart of all such methods. The role of self-consistency was illustrated explicitly in the one-dimensional examples. The most direct way to deal with self-consistency is to compare two geometrical objects by superimposing one on the other; if the two objects can be moved rigidly into perfect coincidence then they are congruent. This is the approach used in classical Euclidean geometry for determining congruence, and it is the approach used in the earlier paper. In this paper, however, we have dealt with the shapes of objects by representing their images in two-dimensional Cartesian coordinate systems, where the mechanisms for comparison are analytical. But self-consistency is still involved in an important way. Recall that in the example of the three-ball plate we used the rigidity of the triangle formed by the three balls to deduce the relative scales, orientations, and origins of the respective coordinate patches. In other words, we derived a Cartesian coordinate system in which the images of the ball plate in each of its placements on the measuring plane were congruent to one another. Having assumed the axiom that we are operating in a euclidean space, the process of derivation itself was tantamount to demonstrating that any other shape of the ball plate would have been inconsistent with the given measurements and measurement procedures. The reason for this is that the euclidean axiom asserts that the measured image of an object measured in one Cartesian reference frame must have the same shape as the image measured in any other Cartesian reference frame. The use of Cartesian coordinates is convenient because it places the arguments in a definite and familiar setting, and it leads to constructive procedures, but their use may obscure the role of self-consistency. Similar trade-offs are common in applications of mathematics. For example, vectors and tensors afford coordinate-free methods that can reveal the underlying nature of a physical process, whereas the use of a specific coordinate system, while possibly more practical, can obscure the underlying process.

A simple exercise involving balls arranged at the vertices of a square lattice is suggested below to show that by rotation about two distinct pivot points, it is possible to arrange that all the balls will fall within a single orbit. In other words, rotations about as few as two pivot points can yield transitive procedures. Although we did not demonstrate procedures for conforming coordinate patches that lie in an orbit generated by rotations about more than one pivot point, it should be clear that the methods demonstrated in the paper extend naturally to such cases. As shown in the earlier paper, such simple rotation procedures can be quite practical. Although practical techniques were not the immediate point of this paper, it is easy to see how careful attention to errors could allow one to translate the idealizations employed here into practical procedures. The practical implications will be examined in a subsequent paper.

I plan to discuss these and related ideas more fully in later work. This paper is the first part of a projected series. Part II will treat self-calibration from the perspective of rigid-motion groups and provide more insight into the relationship between the theories developed here and in Raugh (1985). Orbits and transitivity will reappear there in the mathematical context of *groups acting on sefs* (Armstrong, 1988; Gilbert, 1976; Jacobsen, 1974). Part III will discuss practical implementations and empirical results, and Part IV will discuss calibration in three dimensions.

Exercises. Finally, I would like to suggest some exercises that extend the theory developed in this paper and that allow the reader to test his or her grasp of the concepts.

1) We began with an assumption that each ball of our ball plate was perfectly spherical, but we disallowed knowledge of the radius of each one -- leaving determination of relative radii to the measurement process. It is interesting to note that if the radii of the balls had been known, then there would be no ambiguity of scale. The scales of the various orbital coordinate systems would be well-defined with respect to one another. I leave it to the reader to show that in that case, we would at least know the relative size of each wheel-within-a-wheel, but the relative angular orientations between orbits would remain indeterminate.

2) Go back to the example of the three-ball plate, and note the exercise using a two-ball plate suggested in the footnote. The two-ball plate can be simplified even further. Instead of a two-ball plate, consider a *two-point grid*, i.e., two points marked on the surface of a flat plate. A modified exercise is to show how to use the two-point grid to calibrate two coordinate patches. Don't assume that the rotational center is necessarily on the line connecting the two points.

3) Consider a square ball plate in which the balls lie approximately on the vertices of a square lattice centered and aligned parallel to the edges of the ball plate. The ball plate can contain any number of balls, so long as the number is a square integer. Suppose that two pivot points are selected, one each at the centers of adjacent squares located near the center of the plate. Now let the plate be measured in its first position in the Coordinate Measuring Machine. For the second set of measurements, rotate the plate counterclockwise through ninety degrees about one of the pivot points and measure again. Finally, rotate the plate counterclockwise ninety degrees

about the other pivot point. It is an interesting exercise to show that this set of procedures is transitive, i.e., it gives rise to exactly one orbit. Prescribe procedures for calibrating a Coordinate Measuring Machine using fine and gross rotations around two distinct centers. Note that if the ball plate is radially symmetric about the first center, then it is not quite radially symmetric about the second center, and some of the balls on the "outside edge" will swing beyond the extant coordinate patches. How does this effect the calibration process?

4) Rigid motion of an object, in this paper, refers to continuous motions, and, hence it excludes isometries of the object obtained by reflection. *Reversal* in one and two dimensions refers to such a reflection. It can be thought of as flipping the object over. For example, reversal of a two-dimensional artifact means placing the artifact bottoms-up on the measuring plane and measuring it from above? To illustrate reversal in self-calibration procedures, suppose that we have an artifact marked with a radially symmetric pattern of points similar to the ball plates considered in this paper, and that the artifact is transparent in the sense that the measuring instrument can "see" the pattern through the bottom of the artifact. Now, imagine measuring the artifact right-side-up using fine and gross rotations about a single center of rotation, as described in the paper. Suppose that these procedures give rise to more than one orbit of coordinate patches and that each orbit has been conformed by techniques like the ones described in the paper. Next, reverse the artifact on the measuring plane so that each mark of the pattern falls within a coordinate patch of the same orbit (but not necessarily within the same coordinate patch) as before, when right-side-up, and so that the center of rotation is positioned as before. Finally, measure the reversed pattern in that position. Show that the information obtained from the reversal can be used to determine the relative orientation angles among the various orbits. What does this imply about the definition of transitivity? Can reversal be used to determine the relative scales among various orbits?

5) Show how to combine reversal, as illustrated in the previous exercise, with rotation about a second center of rotation to completely calibrate a square grid pattern like the one in Exercise 3.

6) Suppose that our Coordinate Measuring Machine operated in three dimensions and that we have somehow constructed a "cubical" three-dimensional ball plate, such that the balls are deployed on the vertices of a cubical lattice. Corresponding to the three-dimensional array of balls is a family of coordinate patches lying within the three-dimensional volume of the measurement domain of the Coordinate Measuring Machine. Show that

---

[18] Another way to think of reversal is to imagine that the object is represented by a pattern of points in a Cartesian coordinate system. Reversal of the pattern occurs when the coordinates of all the points are multiplied by an orthonormal matrix of negative determinant.

transitive procedures can be defined for these coordinate patches using just two pivot points.

Acknowledgements

I would like to thank Fabian Pease and Michael Flynn of Stanford University and Geraint Owen of Hewlett-Packard Laboratories for encouraging me to continue work in self-calibration. Robert Hocken of the University of North Carolina at Charlotte and Daniel DeBra of Stanford University both suggested that I discuss the theory of self-calibration in the context of ball plates and Coordinate Measuring Machines to help narrow the gap between classical calibration techniques and the self-calibration techniques of Raugh (1985). I am grateful to them all for their interest and hope that they find something useful in this effort. Louis Jaeckel of RIACS and Hans Samelson of Stanford gave careful readings of early drafts, correcting several errors and suggesting improvements. They are not responsible for any remaining obscurity or error. My wife Vicki Johnson constructed the figures and suggested improvements of the text. I also want to thank her for the use of her Macintosh in preparing the manuscript.

Glossary

Most of the terms that are italicized in the text are gathered here with definitions for easy reference. Those marked by an asterisk are ones that I have used in a novel way, most of which have been adapted from mathematics. Cross-references are indicated by italics within the definitions. I will be interested in comments concerning the appropriateness and utility of these terms.

artifact. Any special object that has been constructed for the purpose of being measured, as for example a *calibration* artifact constructed for the purpose of calibrating a measuring instrument. In two-dimensional calibration these artifacts are usually planar. *Grid plates* and *ball plates* are examples. However, strictly speaking, a ball plate is a three-dimensional object, but when estimating point positions we assume that the **XYZ** coordinates can be projected without significant loss of accuracy onto the XY plane.

ball plate. A rigid artifact used for calibrating two-dimensional coordinate-measuring instruments. It consists of a flat plate with an array of small spheres mounted on short pedestals on the upper surface at more-or-less the same height. The balls may be of equal or unequal radii and may be arranged on the ball plate in a regular or irregular pattern. Ball plates are measured by a *Coordinate Measuring Machine.* (Hocken and Borchardt, 1979; Reeve, 1974.)

calibrate. To assign physically meaningful and correct values to the graduations of a measuring device, as, for example, a ruler or thermometer. For an example in two dimension, the digital readouts of the interferometers that track the motions of a measurement probe or the stage of a two-dimensional measuring device, such as a *Coordinate Measuring Machine* or an electron-beam lithography system, respectively, may be thought of as the "graduations" of the measuring device. In the text, such two-dimensional graduations are referred to as *coordinate markers.* Although these readouts may constitute an approximately *orthonormal coordinate system,* they are usually not accurate enough, and they contain unspecified curvilinear components. To calibrate such a system of coordinate markers means to determine a transformation of the coordinate markers to a physically correct system of orthonormal coordinates. The transformation so determined is referred to as a *calibration,* a *calibration function,* or a *calibration mapping.* Note that, in this paper, a calibration mapping is not unique, since we do not require specification of scale, orientation, or origin of the orthonormal coordinate system. The term calibrate may also apply to the measurement of an object to be used as a *measurement standard,* as, for example, we may say that a calibration *artifact* has itself been calibrated, i.e. measured with sufficient accuracy to be used for calibrating an instrument. In the text, examples are given of calibration procedures for a limited number of *coordinate patches* in the plane of a two-dimensional measuring instrument, In such a case, calibration only refers to the relevant subset of the plane. See *self-cal ibra te.*

Cartesian coordinate system. In this paper, synonym of *orthonormal coordinate system.* Cartesian coordinates are the kind used in classical calculus and analytic geometry, namely the Cartesian method for handling euclidean geometry in a systematically algebraic and numerical way. For two-dimensional geometry, the coordinate plane is represented by two coordinate axes set at right angles to each other, lengths are represented linearly on each axis, and the axes are scaled equally. The Pythagorean theorem is used for computing distances between points, and the methods of plane trigonometry are use to derive formulas for angles between lines. Although these mathematical procedures are simple and well-established, it is a profound question whether euclidean geometry is a correct description of the physical universe. In the idealizations of this paper euclidean geometry is accepted as axiomatic, and it is assumed that point positions can be measured with perfect accuracy and computed without roundoff error. For reference, see *orthonormal coordinate system.*

calibration, calibration f-unction, calibration mapping. See *calibrate, classical approach to calibration, self-calibrate.*

classical approach to calibration. A technique in which a calibrated *artifact,* such as a *standard calibration grid,* is used to calibrate a measuring instrument. The idea is straightforward. The artifact is measured and **a** transformation is derived to convert the machine coordinates to the coordinate system of the artifact. I refer to calibration problems that are solved by classical means as standard calibration problems. Compare with *self-calibration.*

CMM. *Coordinate Measuring Machine.*

conformed coordinate systems.* A way of saying that a given set of coordinate systems within a plane are identical. For example consider a set of *coordinate patches* in the *measuring plane* of a measuring instrument. Suppose, as in the text, that each coordinate patch has been rendered as an *orthonormal coordinate system,* but no relationship has been given among the scales, relative angles of orientation, or placement of the origins. In this situation these coordinate patches are not *conformed.* However, if by some means, such as by measurement strategies exemplified in the text, it were possible to transform all of the coordinate systems so that they shared a common origin, were all of the same scale, and the respective axes **were** aligned, we would say that the coordinate patches were conformed. *Calibration* of a two-dimensional measuring instrument may be viewed **as a**

set of procedures that enable the coordinate patches of the instrument to be conformed to one another. Although it may not be obvious, it is important to note that, in the text, the process of conforming a set of coordinate patches is carried out as a means of achieving *self-consistency* for a given set of *measurement procedures.*

congruent. A term from plain geometry for saying that two objects are of identical size and shape. Two geometric figures are said to be congruent if the one can be moved rigidly into perfect coincidence with the other. Two figures are congruent if they are *similar* to one another and of equal size. Compare with *shape.*

coordinate markers.* A convenient synonym for coordinates that may not have any physical significance or may not be adequately accurate. For example, if arbitrary graduations were notched on a thermometer or a yardstick, they would be called coordinate markers. If then physically correct numbers were attached to each of the markers, for example, degrees Celsius or lines and inches, respectively, we would call the attached set of numbers coordinates. The act of assigning physically meaningful values to coordinate markers on a measuring device is called *calibration*. In the case of two-dimensional measuring instruments, as for example *Coordinate Measuring Machines*, putative coordinates are given as readouts of a split-beam interferometer that tracks approximately orthogonal components of the motion of a probe; typically, however, the coordinates are not reliably *orthonormal* or accurate. To underscore the fact that such coordinates have unknown systematic error, they may be called coordinate markers. To calibrate such a system of coordinate markers, one must associate or "map" the coordinate markers to physically meaningful (and correct) numerical values. The *machine coordinate space* is a convenient means of depicting coordinate markers geometrically, and the *machine calibration space* is a convenient way of representing calibrated coordinates.

Coordinate Measuring Machine (CMM). A two-dimensional measuring device consisting of a plane on which an *artifact,* such as a *ball plate* can be mounted and measured. The measurement is done by moving a probe to the point of interest. X-Y-Z translations of the probe are measured interferometrically. Although a Z component does exist for small vertical displacements, in this paper we are concerned only with the projections onto the X-Y plane.

coordinate patch.* A two-dimensional coordinate system defined for a small neighborhood of a larger surface, used to refer to small subsets of the *measurement plane* of a two-dimensional measuring device. Although the coordinate system of, say, a *Coordinate Measuring Machine* may have systematic curvilinear errors when viewed on the macroscopic scale, we assume that in small neighborhoods the coordinate system is linear to first order. By concentrating attention on coordinate patches, we develop procedures for transforming the linear coordinates of each patch into an *orthonormal coordinate system,* respectively and for *conforming* the various orthonormal coordinate systems into one common orthonormal coordinate system. The procedures involve measuring a *calibration artifact* in various placements on the measuring plane of the two-dimensional measuring device. The process of conforming the various coordinate patches, then, is aimed at producing *self-consistency* in the measurements of the artifact. If the *measurement procedures* are *determinate,* the resulting conformed coordinate system constitutes a *calibration* of the Coordinate Measuring Machine. Note that in the examples in the text, the coordinate patches under consideration usually constitute only a small subset of the entire *measurement plane,* nevertheless when all of such coordinate patches are conformed we still call it a calibration.

determinate measurement procedures.* In measuring a *pattern* **on an** uncalibrated measuring device, the only observable product of the *measurement procedures* is the *coordinate markers* obtained for the points of the pattern, or transformations thereof. These coordinate markers can be plotted as points of an *orthonormal Cartesian coordinate system,* called the machine coordinate space. If the pattern is measured in each of several placements of the pattern on the *measuring plane,* then the several *images* of the pattern can be depicted in the *machine coordinate space* or in *machine calibration space.* These images are observable, and their shapes may be compared. If all of the shapes are *congruent* to one another, then the measurement procedures are said to be *self-consistent.* Under certain conditions, some of which are specified in the text, self-consistent measurements can be used to deduce the correct shape of the pattern and, simultaneously, to derive a *calibration* of the measuring device. In such cases, the measurement procedures are said to be *determinate.* Under other conditions, it may not be possible to verify the shape of the pattern or to derive a calibration of the device, as is the case for *non t ransit ive measurement procedures.* In the latter case, the measurement procedures are said to be *indeterminate.*

equivalence class. A term used in mathematics to denote a partition of all the members of a set, i.e., a division of the set into a family of subsets such that each member of the set belongs to one, and only one, subset. Each such subset is called an equivalence class. *Orbits* are equivalence classes.

feature. A generic term referring to any measurable part of an *artifact* or rigid object that is to be measured by a measuring device. For examples, if the object is a *grid plate,* then the features are grid points; if the object is a *ball plate,* then the features are the balls or the centers of the balls inferred from measurements of the surfaces.

fine rotation.* **A** very small rotation of an *artifact,* such as a rigid two-dimensional *pattern* or a *ball plate,* about a fixed *pivot point* such that each pattern point or ball, respectively, stays within a single *coordinate patch.* Distinguished from *gross rotation.* The ball plate or pattern is assumed to be mounted on the *measurement plane* of a two-dimensional measuring device, and each ball or pattern point is assumed to lie within a coordinate patch. Note that fine rotations move each pattern point or ball on the arc of a circle.

fixpoint.* Synonym for center of rotation, rotational fixpoint, or *pivot point.*

grid plate. A planar *artifact* containing a *pattern* of measurable points embedded in its surface. The pattern can be arbitrary, as for example a regular array of points such as a square lattice or a rotationally symmetric array, or it can be a random pattern sprinkled evenly or unevenly like freckles.

gross rotation.* A large rotation that moves each *feature* of interest of an *artifact* out of one *coordinate patch* and into another. Distinguished from fine rotation.

image of a measured object. Refers to a representation of an object in a suitable coordinate system. In the paper, measurements of the *features* of a two-dimensional object are given as a pair of coordinates, which are then plotted in a *Cartesian coordinate system,* such as *machine coordinate space* or *machine calibration space.* Such a plot is referred to as an image of the object.

indeterminate measurement procedures.* *Measurement procedures* that are not *determinate* are called indeterminate. Here is a way to test whether measurement procedures for a specific machine and pattern of points are indeterminate: Examine whether, theoretically, the same measurement procedures carried out on a different measuring device, using a *different-shaped pattern* of points, can yield exactly the same set of measurements as in the case under question. If so, the measurement procedures are indeterminate.

internal angles. A way of representing the *shape* of an object. The internal angles of an object are the angles formed between every set of three points within the object. Two objects with identical corresponding angles have the same shape and are said to be *similar.* If the two object are also of identical size, they are said to be *congruent.*

machine calibration space.” A geometric way to represent a transformation of *machine coordinate space* to test the transformation for *self-consistency.* Consider a two-dimensional measuring instrument, such as a *Coordinate Measuring Machine.* The machine coordinate space provides a simple means of visualizing measured objects as geometric figures. If the measuring instrument is *uncalibrated,* then the image of an object in machine coordinate space will not be of the same *shape* as the object itself, but at least the image provides something observable. Moreover, if the object is measured in several placements on the *measuring plane* of the instrument, it is likely that the resulting *images* will not be *congruent* to one another. Now consider a transformation of the machine coordinates to some other *Cartesian reference frame* (a Cartesian frame different than the machine coordinate space). *Self-calibration* **involves** the study of transformations that transform the dissimilar images of the machine coordinate space into congruent images in the latter Cartesian framework, called the machine calibration space. Its importance arises from the fact that when a given *pattern* is measured in various placements on the measuring plane of the measuring device, we can compare and alter the shapes of the images in the machine calibration space, even though we cannot observe the shape of the pattern itself. When a transformation of machine coordinates is found for which the images are all congruent to one another (i.e., of equal size and shape), we say that the coordinate transformation is *self-consistent.* **As** shorthand, when it is clear that we are applying a self-consistent transformation we will say the the *measurement procedures* themselves are self-consis tent. The theory of self-calibration is concerned with specifying conditions under which self-consistency is necessary or sufficient for determining a calibration of the measuring machine. *Transitivity* of self-consistent measurement procedures is a necessary condition but is not sufficient.

machine coordinate space.* **A** simple geometric way to represent the *coordinate markers* of a measuring device. Consider the example of a two-dimensional measuring instrument, such as a *Coordinate Measuring Machine.* Although the coordinate system of the device may be *uncalibrated,* it can still be represented as points of a two-dimensional *orthonormal coordinate system.* In the text, we envision measuring a *pattern* of points to obtain machine coordinates for each point and plotting the points on Cartesian graph paper, using the machine coordinates for each point. We call this Cartesian framework the machine coordinate space. If the machine is uncalibrated, then the *image* of the pattern on the graph paper will not be the same shape as the original pattern. The importance of this representational technique arises from the fact that when a given pattern is measured in various placements on the *measuring plane* of the measuring device, we can compare the shapes of the images in the machine coordinate space, even though we cannot observe the shape of the pattern itself. If by chance the shapes were all *congruent,* we would then say that the *measurement procedures* were *self-consistent.* The theory of self-calibration involves analysis of transformations of machine coordinates that yield congruent images.

measurement. In this paper, the determination of the *shape* of an object by finding coordinates in a single *orthonormal* reference frame for all the *features* of the object. We are not concerned here with the absolute size of an object but are only concerned with its shape. Note that the shape of an object may be equivalently given by specifying the *internal angles* among all pairs of features. A method used in the text to show that *measurement procedures* are *indeterminate,* is to show that two objects with distinctly different internal angles are consistent with the given measurements and measurement procedures. In the same way that in measuring an object we are not concerned with the size of the object but only its shape, in *calibrating* a measuring device we are not concerned with the scale, orientation, or origin of the orthonormal coordinate system that we assign to the measuring device but are only concerned that the measuring device give the correct shape of any measured object. It is useful to consider the duality of measurement and calibration: An object can be measured by a calibrated measuring device, and a measuring device may be calibrated by comparison to a *measurement standard.* It is the mutual dependency of calibration and measurement that gives the problem of self-calibration such an interesting chicken-or-egg quality. The challenge of self-calibration is to devise procedures that permit simultaneous calibration of a measuring instrument and measurement of the calibration *artifact.*

measuring plane. That part of a two-dimensional coordinate measuring device, for example a *Coordinate Measuring Machine,* on which an *artifact* can be fastened for measurement.

measurement procedures. A shorthand way of referring to the major ingredients of a measuring experiment, namely, the *artifact* to be measured, the measuring device, the series of placements of the artifact on the *measuring plane* of the measuring device, and the transformation from machine coordinate markers to *machine calibration space.* Thus, *transitivity* and *self-consistency* are defined with specific reference to the measurement procedures, meaning that all the named elements are involved.

measurement standard. An accurately measured *artifact* used for *calibrating* a measuring device. The National Institute of Standards and Technology (NIST) provides such measurement standards for a wide variety of applications. The need for *self-calibration* arises because neither the NIST nor any other institution has been able to provide an accepted two-dimensional measurement standard that is accurate to within ten nanometers, the magnitude of precision that current metrology tools are capable of.

nontransitivity? Measurement procedures that are not *transitive* are said to be nontransi tive.

orbit? *A* series of *coordinate patches* that are related to one another through the rigid motion(s) of the *features* of a *calibration artifact.* When *gross* rotations of a rotationally symmetric rigid object are made, such as a *ball plate* or *grid pattern,* on the *measurement plane* of a two-dimensional measuring device, each feature of the object (i.e., ball or pattern point) moves from one coordinate patch into another, while the feature that had previously occupied the new coordinate patch has itself moved on to still another, and so on. Thus each feature pursues another feature, and it is also pursued by some feature. If you start with any feature (on an artifact containing only a finite number of features) and connect it to the one it pursues, and then connect that feature to the one it pursues, and so on, you must eventually arrive back at the feature you started with. Such a chain of features may or may not exhaust all the features of the grid pattern or ball plate. In any case, the set of coordinate patches traversed by such a chain of features is called an orbit. Each feature belongs to one and only one orbit. It can happen that either one orbit embraces all of the coordinate patches, in which case the *measurement procedures* are called *transitive,* or there are more than one orbit, in which case I call the procedure *nontransitive.* When there are multiple orbits they are all disjoint, and their union comprises all of the coordinate patches. Orbits are important in *self-calibration* because only features within a single orbit can be measured with respect to one another. Thus the *shape* of a *pattern* of points, each of which falls within the same orbit, can often be deduced from the measurements of an uncalibrated measuring device. But the relative orientation and scale of patterns in distinct orbits cannot be determined.

orthonormal coordinate system. In this paper, a synonym of *Cartesian coordinate system.* A system of orthogonal axes in which the coordinates are marked off linearly on each axis and the scales of all the axes are identical. It provides a standard framework for dealing with euclidean geometry analytically. (See any introductory book on calculus and analytic geometry or linear algebra; for a thorough treatment from the standpoint of finite-dimensional vector spaces see Efimov and Rozendorn, 1975.)

pattern. An array of points situated on the surface of a rigid plate or other *artifact.* A pattern can be regular, as in a square lattice of points, or it can consist of a random array, like freckles.

pivot point.* Synonym for center of rotation or *fixpoint.*

primal position.* The first position a *feature* is in when the *artifact* is first place on the *measuring plane* of a measuring instrument, or the first *coordinate patch* in which a feature lies.

principle of self-calibration.* *Self-calibration* is possible only when *transitive measurement procedures* are used. Transitivity and *self-consistency* are necessary but not sufficient conditions for self-calibration. I consider this to be the fundamental fact about self-calibration.

self-calibration.* A type of *calibration* in which neither the measuring instrument to be calibrated nor the calibration *artifact* is calibrated. In this case, the instrument and the object must be calibrated and measured simultaneously. The problem is more difficult than *classical calibration problems.*

self-consistency.* When the images of a *pattern* of points in various specific placements on the *measuring plane* of a measurement device are all *congruent* to one another (i.e., of equal size and *shape)* in the *machine coordinate space or the machine calibration space, we* say that the *measurement procedures* are self-consistent. A primary aim of *self-calibration* procedures is to produce self-consistency.

shape. The shape of an object as opposed to its size. In this paper we are concerned with determining the shapes of objects and are not concerned with size *per se.* The shape of an object can be determined by either representing it in a well-defined coordinate system, such as *Cartesian coordinates,* or by the *internal angles* of the object. Objects of the same shape are said to be *similar.* Objects of the same size and shape are said to be *congruent.*

similarity. A term from plain geometry indicating comparison of two objects. Two triangles are said to be similar if their corresponding angles are equal. By extension, two geometric figures are said to be equal if all corresponding triplets of points in the two figures form similar triangles. Similarity thus refers to the *shapes* of objects and not their relative sizes.

two-point grid. A flat *calibration artifact,* consisting of two measurable points on the upper surface. This is a theoretical construct that can be used to illustrate the basic operations used in this paper for calibrating a two-dimensional measuring machine. A use for the two-ball plate is proposed as an exercise in the conclusion of the paper as a means of testing the reader's understanding of some of the basic ideas.

transitivity.* Applied to *measurement procedures* that give rise to a single *orbit.* Transitivity, along with *self-consistency,* is a necessary but not sufficient condition for *self-cal ibra t ion.* Transitivity and self-consistency are the two key concepts in self-calibration.

logical principle of transitivity.* Two quantities equal to a third are equal to each other. This principle, or axiom of logic, underlies the theory of *self-calibration. Cal ibra t ion* ultimately involves direct or indirect comparison of all *features* of an *artifact,* and the same is true for all the parts of the coordinate system of a measuring machine (i.e., coordinate patches), for they all must be compared to one another.

wheels-within-wheels.* A descriptive expression used in the text to describe the kind of *indeterminacy* that exists for *nontransitive measurement procedures* resulting from rotation of the measurement artifact around a single *pivot point.* Each *orbit* of a point *pattern* that is radially symmetric about the pivot point can be *calibrated,* but the relative scales, orientation angles, and placement of the origins among disjoint orbits must remain indeterminate. In fact, as explained in the text, the relative orientations and scales are arbitrary, meaning that any rotation of the orbits with respect to each other, and any scale relationships between them, is consistent with the measurements. What this means is that the *calibration artifact* can be illustrated by "wheels-within-wheels," such that the *shape* of all the *features* falling within any given orbit is well determined, but each such ring of features is free to expand or rotate arbitrarily around the pivot point.

References

Armstrong, M. A., *Groups and Symmetry,* Springer-Verlag, New York, 1988.

Buck, R. Creighton, *Advanced Calculus,* McGraw-Hill, New York, second edition, 1965.

Efimov, N. V., and E. R. Rozendorn, *Linear Algebra and Multi-Dimensional Geometry,* MIR Publishers, Moscow, 1975.

Gilbert, William J., *Modern Algebra with Applications,* Wiley-Interscience, New York, 1976.

Halmos, Paul R., *Finite-Dimensional Vector Spaces,* Springer-Verlag, New York, 1974

Hocken, R. J. and B. R. Borchardt, On Characterizing Measuring Machine Geometry, National Bureau of Standards Report NBSIR 79-1752, Washington, D. C., May 1979.

Jacobsen, Nathan, *Basic Algebra* I, W. H. Freeman and Company, San Francisco, 1974.

Petersen, Karl, *Ergodic Theory,* Cambridge University Press, New York, 1983.

Protter, Murray H. and Charles B. Morrey, Jr., *Modern Mathematical Analysis,* Addison-Wesley Publishing Co., Reading, MA, 1964.

Raugh, M. R., "Absolute two-dimensional sub-micron metrology for electron beam lithography: A theory of calibration with applications," Precision Engineering, 7(1), pp. 3-13 (Jan 1985).

Reeve, Charles P., A Method of Calibrating Two-dimensional Reference Plates, National Bureau of Standards Report NBSIR 74-532, Washington, D. C., July 26, 1974.

Segel, Lee A. and G. H. Handelman, *Mathematics Applied to Continuum Mechanics*, Macmillan, New York, 1977.

Sokolnikoff, I. S., *Mathematical Theory* of *Elasticity,* McGraw-Hill, New York, 1956.