

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 614

6 June 1981

**Equation Counting and the Interpretation
of Sensory Data**

W.A. Richards, J.M. Rubin & D.D. Hoffman

ABSTRACT

Many problems in biological information processing require the solution to a complex system of equations in many unknown variables. An equation-counting procedure is described for determining whether such a system of equations will indeed have a unique solution, and under what conditions the solution should be interpreted as "correct". Three examples of the procedure are given for illustration, one for auditory signal processing and two from vision.

Acknowledgement: This report describes research in Natural Computation at the Artificial Intelligence Laboratory and Psychology Department of the Massachusetts Institute of Technology. Support for this research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505 and by NSF and AFOSR under grant 7923110-MCS. D.D. Hoffman was also supported by a Hughes Aircraft Company graduate fellowship and J.M. Rubin was supported by an NSF Fellowship. We greatly appreciate the help of Professor M. Artin and T. Shifrin, who pointed us to the Bezout Theorem and the Jacobian test, and also the insightful comments of Professor S. Ullman and A. Pentland. Technical assistance in the preparation of this manuscript was provided by C.J. Papineau.

© MASSACHUSETTS INSTITUTE OF TECHNOLOGY 1981

1. Introduction

Sensory data are routinely interpreted as external events by biological systems. This achievement is the classical problem of perception: given a pattern of sensory activity, what are the external events that caused this activity? In order for an organism to survive, such assignments of cause, or interpretations, must be reliable and appropriate. Yet the sensory data by themselves are ambiguous (as illustrated by the projection of the three-dimensional world onto our two-dimensional retina). The appropriate interpretation of a pattern of activity is thus just one of many possibilities. The objective of this paper is to outline the power and pitfalls of an equation-counting procedure, and how this procedure can lend insight into the interpretation process.

The ambiguity of the sensory activity becomes very clear when formal relations are developed between these sense data (the givens or "knowns") and the external events (or "unknowns") that generate the data (Marr, 1976, 1982; Ullman, 1979). When such relations are expressed in the form of equations relating the "knowns" to the "unknowns", then the number of unknowns will almost always exceed the number of equations. The incompleteness of the set of equations is a consequence of the fact that the mapping of a world event into the sensor entails a loss of information and hence is usually many-to-one. But if the system of equations is incomplete, with the number of equations less than the number of unknowns, then the system cannot be solved uniquely and constructing a unique description of the external event becomes impossible.

Fortunately, events in the real world are not arbitrary, but are constrained by natural laws. The sense data reflect these constraints (Huffman, 1971; Clowes, 1971; Waltz, 1975). Once discovered, these additional relationships can yield the remaining equations needed to make the number of equations equal to the number of unknowns. A unique solution to the set of equations may then be sought, permitting an interpretation of the data. (The correctness or validity of the interpretation will be discussed later.)

The paper begins with a rather simple example of "equation-counting," namely, the detection of a narrow-band signal in noise. This problem involves only linear equations, but still illustrates the general features of the approach and raises three issues: 1) independence of the equations; 2) constraints needed to yield a unique solution, and 3) whether this unique solution is indeed "correct". We then introduce a theorem by Bezout which is needed to place bounds on the number of possible solutions to polynomial equations, as well as a Jacobian test for the independence of these equations. Finally, two other problem examples are given to illustrate further details. One example concerns recovering structure from visual motion; the other shows why three spectral samples are needed to distinguish shadows from reflectance changes.

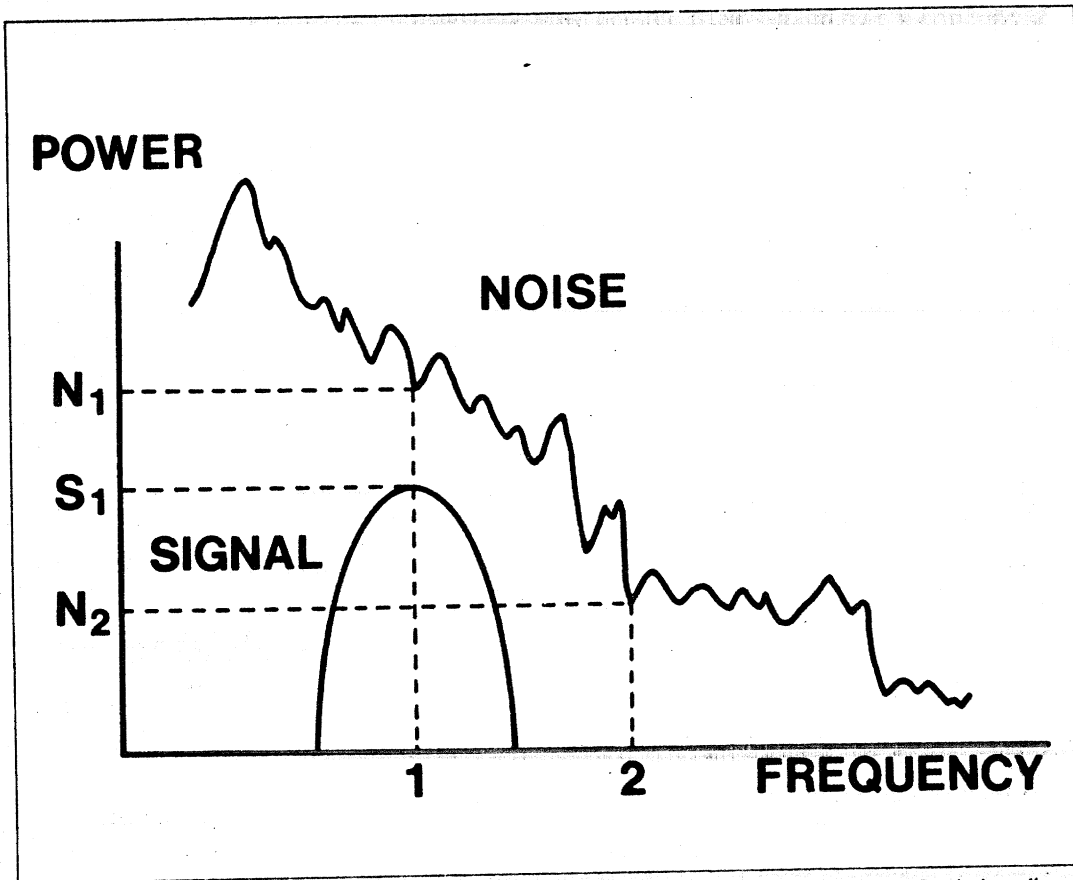


Figure 1. An illustration of a narrow band signal against a background of noise. The noise is broadband with a constant time averaged spectrum.

2. A Classic Problem

A problem faced by many animals is the need to isolate a narrow-band, species-specific signal from the background noise. Although examples may be found in every sense modality, the clearest probably occur in audition. Consider the bird listening to the call of its mate in the forest of other sounds; the dog perking his ears at his master's whistle; or the moth's task of isolating the cry of the bat as it homes in for its next meal. In each case, the signal is confined to a relatively narrow band, as illustrated in Figure 1, whereas the competing noise is much broader. Given that the frequency band of the signal is known (as it would be for the bird or the moth), how many intensity samples must be taken to isolate the signal from the noise?

Clearly, by referring to Figure 1, we see that sampling in the signal-band at frequency 1 will not allow us to isolate the signal. More formally, the ear will receive intensity I , at frequency f_1 equal to the sum of the power produced by each source:

$$I(f_1) = S(f_1) + N(f_1) \quad (1)$$

where S corresponds to the power of the narrow-band signal at f_1 and N is the background noise at the same frequency. Since only I is available to the listener, S and N cannot be separated, for we

have only one equation in two unknowns, S and N . More generally, if we allow additional samples at time intervals t_j , then equation (1) can be generalized to:

$$I(f_1, t_j) = S(f_1, t_j) + N(f_1, t_j) \quad (2)$$

Thus, for T time samples we will obtain T equations in $2T$ unknowns, which will not permit a unique solution for S .

Let us now make the obvious next step and consider frequency samples outside the signal band. The frequency f_1 in equation (2) then becomes indexed to f_i . However, since the signal is zero outside the band at f_1 , then $S(f_i, t_j) = 0$ for $i \neq 1$. These conditions may be expressed as two equations:

$$I(f_i, t_j) = S(f_i, t_j) + N(f_i, t_j) \quad (3a)$$

$$S(f_i, t_j) = 0, \quad (i \neq 1) \quad (3b)$$

Letting F and T be the number of frequency and time samples, respectively, there will be a total of $F \cdot T$ equations of form (3a) and $(F - 1) \cdot T$ equations of form (3b). The total number of equations is thus $2 \cdot FT - T$. Similarly, the total number of unknowns will be $F \cdot T$ for S and $F \cdot T$ for N or $2 \cdot F \cdot T$. In order to solve uniquely for S , the minimum condition is that the number of equations E equal (or exceed) the number of unknowns U :

$$E \geq U \quad (4)$$

For solution, equations (3a,b) thus must pass the following inequality test:

$$2FT - T \geq 2FT \quad (5)$$

or

$$0 \geq T$$

which fails since $T \geq 1$. Thus a narrow-band signal cannot be extracted from the broad-band noise without specifying further constraints upon either the signal or the noise.

2.1 Flat Noise Condition

Very often noise is relatively constant over frequency (or time), for example, the hum of an air conditioner, a steady wind flow passing the body, or even body noise. This condition can be expressed by the following relation:

$$N(f_i, t_j) = N(f_i, t_j) \quad (6)$$

where $i \neq 1$ and f_1 serves as the reference frequency. We now see that for a total of F frequency samples, equation (6) adds $(F - 1) \cdot T$ equations but no more unknowns. Applying the Inequality Test (4), we now find:

$$(2FT - T) + T(F - 1) \geq 2FT \quad (7)$$

or

$$F \cdot T \geq 2T$$

or

$$F \geq 2 \quad (8)$$

Thus, the minimum condition for a unique solution occurs for two frequency samples at any temporal interval. Ignoring the time variable, equations (3a, b), and (6) then become

$$I(f_1) = S(f_1) + N(f_1) \quad (9)$$

$$I(f_2) = S(f_2) + N(f_2)$$

$$S(f_2) = 0$$

$$N(f_1) = N(f_2)$$

We now have four equations in four unknowns, which allows us to solve for $S(f_1)$, given that the noise spectrum is flat.

2.2 Independence and Uniqueness

Although two frequency samples plus the constraint of "flat noise" yield the same number of equations as unknowns, these equations must be shown to be independent. Certainly we can reduce equations (9) to obtain an explicit solution for $S(f_1)$, thereby demonstrating independence. However in the more complex cases normally encountered, such a reduction is often difficult or may be impossible (for example if fifth degree polynomials are involved). We therefore seek a more general test for independence.

In the above example, the obvious test is to recast equations (9) so all the unknowns are on the right hand side (R.H.S.) of the equality, and all the knowns are in the L.H.S. Then the determinant of the coefficients of the R.H.S. can be calculated. By "Cramer's Rule", we know that if this determinant is not zero, then the equations have a unique solution (Thomas, 1951). To proceed, equations (9) are rearranged so the unknowns are ordered in the sequence $S(f_1)$, $N(f_1)$, $S(f_2)$, $N(f_2)$ and are each aligned in their separate columns on the R.H.S. of the equality. Since there are four unknowns and four equations, the matrix of the coefficients of the unknowns will be as follows:

$$\begin{vmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{vmatrix} \quad (10)$$

The determinant of this matrix is easily found to be 1 (i.e., it has maximum rank), and hence the set of equations (9) must have a unique solution.

We now can proceed with confidence to find the following solution for $S(f_1)$:

$$S(f_1) = I(f_1) - I(f_2) \quad (11)$$

2.3 Corroboration and Constraint

Unfortunately, *any pair* of sensory intensities $I(f_1)$ and $I(f_2)$ will provide a value for $S(f_1)$. How do we know, therefore, that the obtained value for $S(f_1)$ is indeed correct? Clearly if the noise stimulus is not flat over frequency, but varies as shown in Fig. 1, then the solution for $S(f_1)$ will be wrong because the assumed condition does not apply. Without some evidence supporting the "flat noise" assumption, a meaningful interpretation of the intensity values $I(f_1)$, $I(f_2)$ cannot be made.

Ideally, any assumed condition, such as the flat noise condition, that is introduced to match the number of equations to the unknowns should be a regularity in the world or a "law" that is never (or rarely) broken by nature. Such conditions are difficult to discover, but when found and introduced into the system of equations provide powerful *constraints* on the solutions. Often the constraint may be a statistical regularity (Witkin, 1980; Pentland, 1980). Poor choices for constraints are those conditions that are very narrow and restrictive and which do not capture a very general property of the world.

In the case of detecting a narrow-band signal in "flat-noise", the imposed condition is very restrictive. However, some attempt can be made to verify the validity of invoking this condition. For example, one possibility might be to examine other frequencies to see if the relation $N(f_1) = N(f_2)$ holds for a range of frequencies outside the signal band. (Note that the solution for $S(f_1)$ should also hold.) If so, then the chance that the "flat-noise" condition is invalid is reduced, although the uncertainty is never eliminated. Sampling at additional frequencies thus provides some (weak) corroboration for the interpretation, increasing its likelihood. (In fact, the condition assumed here has merely been replaced by another, less restrictive assumption about the smoothness of waveforms.) Stronger forms of corroboration will be discussed in later sections.

Finally, it should be noted that in cases where the imposed conditions are not verifiable, the appropriateness of the condition can often be rejected quite easily. For example, if $S(f_1)$ is found to be negative, then since negative signals are not physically realizable, the assumption must not be valid. This strategy of rejecting certain conditions or possible states of the environment has been found useful elsewhere (Rubin and Richards, 1981).

3. Non-Linear (Polynomial) Equations

3.1 Bezout's Theorem

In the above example, all of the equations were linear, and simple techniques of linear algebra could be used. What if one or more of the equations were quadratic or a still higher degree polynomial? In such cases, which are quite common, each n th order polynomial will at most have n distinct roots. How many possible solutions will there be if there are M polynomial equations of degree N ? Can we even guarantee that there will in fact be a finite set of solutions? If this cannot be guaranteed, then the test that states the number of equations E should at least equal the number of unknowns U is not useful, and the simple equation-counting procedure collapses at the onset. Fortunately, Bezout's Theorem tells us under what conditions a finite set of solutions can be found to N equations in N unknowns, and just what the maximum number of solutions will be (Van der Waerden, 1940).

Theorem (Bezout): A set of N independent polynomial equations in N variables will have a maximum number of generic solutions equal to the product of the degrees of the equations.¹

The above theorem is critical for our procedure because it states that if the relations among the N variables can be cast as N independent polynomial equations (perhaps by a change in the form of the variables), then there will be a finite set of isolated solution points. Furthermore, this set will include all the possible solutions. (See Appendix II for a brief discussion of a generalization of Bezout's Theorem by Sard to include any set of smooth functions on manifolds.) For linear equations, it is clear that the product of the degrees of the equations will always be one, and only one solution set will be found. For third order equations, which may include terms such as $x \cdot y \cdot z$, or $y^2 \cdot z$, the number of possible N -tuples of variables that satisfy the N equations can be quite high. Among these is the physically meaningful solution that we seek, provided our hypotheses are correct.

3.2 The Jacobian Test

Bezout's Theorem states that in principle, N polynomial equations of any degree can provide a solution to N unknowns, if the equations are independent. In our simple first example, the deter-

¹By a generic solution, we mean that a slight perturbation in the values of the variables will not alter the solution appreciably (as would be the case if the solution were the special case of two circles just grazing each other rather than intersecting, for example).

minant of the matrix of coefficients of the unknowns was used to check for independence. More generally, the Jacobian of the set of equations should be evaluated (Kendig, 1977; Guillemain and Pollack, 1974). The Jacobian is formed by taking all N partial derivatives of each of the N equations ($\partial f_1/\partial x_1, \partial f_2/\partial x_2, \dots, \partial f_n/\partial x_n$), and placing these partial derivatives in an $N \times N$ matrix, where the columns represent each unknown and the rows correspond to the equations. Clearly, for linear equations, the Jacobian is simply the matrix of the coefficients of the unknowns of each equation.

Jacobian Test (for Independence): If the determinant of the Jacobian of the system of N equations in N unknowns is non-zero, then a countable set of isolated solution points can be found.

This test is simply an application of the Inverse Function Theorem, which gives a condition for a one-to-one and onto mapping between real variables. Note that if the determinant of the Jacobian collapses to zero (by a loss of rank), then this is *not* a proof that solution points cannot be found. The Jacobian test is therefore a test for sufficiency, not necessity.

3.3 Summary of Procedure

To apply the "equation counting" method to the recovery of event descriptions from limited sensory data, we therefore proceed as follows:

1. Set up polynomial equations describing the mapping of the external (unknown) variables into the (known) sense data.
2. Embody as many constraints as necessary in the form of additional polynomial equations relating the variables in order that the total number of equations equals the number of unknowns that are to be recovered. Whenever possible, choose "constraints" that can be verified from the data. Those that capture a regular or consistent property of the world are the best choice.
3. Apply the Jacobian test to demonstrate that the equations are independent. Bezout's Theorem then guarantees that there will be a finite number of solution points. If the Jacobian test fails, try to discover new constraints. (See also Section 5.6.)
4. Proceed to solve for the variables of interest. (We know of no simple heuristics for this step.)
5. Demonstrate that all constraints and conditions are valid. Usually this will involve taking an extra, independent measurement and verifying that the same solution is obtained. Some care must be taken with this step, however, as will be seen in the examples to follow.
6. The sense data may now be given a preliminary interpretation. However, a final interpretation should await two further tests to be described subsequently. One is the exclusion of competing interpretations, the other is corroboration, using an independent system of equations. (See Sections 6.0 and 6.1.)

4. Two Examples

4.1 Example 1: Recovering Structure from Motion

The difference in visual impressions between a static scene and a dynamic movie is often quite striking. Somehow the motion created by viewing a rapid sequence of frames will transform an ambiguous 2-D shape into a vivid 3-D structure. Perhaps the most common example of this phenomenon occurs when we walk, run, or drive and immediately *know* the spatial configuration of the objects about us, regardless whether we use two eyes or one. Although Ullman (1979) has shown how the spatial relations may be recovered using motion information in the general case, we wish to consider a simpler version of the same problem that has a more compact solution: namely, given a person in locomotion, how can he recover the orientation of the surface on which he walks?

Let the surface be covered with markings, or for convenience, let a short "stick" lie on the surface patch of particular interest. Then if the observer looks at the center of the "stick" as he moves ahead, the image of the "stick" as seen on his retina will rotate and change length as shown in frames $F1$, $F2$, and $F3$ of Fig. 2. Because the stick lies in a plane of fixed orientation relative to the moving observer, the orientation of the surface patch can be specified by the axis of rotation of the "stick". The problem then is equivalent to recovering the axis of rotation of a rotating rod seen by a stationary observer.

Figure 2 illustrates the general form of this common problem. The "stick" or rod is rotating in 3-space and is projected onto a single 2-D retina. Let each of these retinal images be discrete time samples or frames as in a TV. Given only the three (or more) ambiguous 2-D image frames $F1$, $F2$, $F3$, how can the axis of rotation of the rod be recovered? This is a task that is solved easily by the human observer, although no information other than the 2-D motion of the end points of the rod is available (Johansson, 1975).

The inset to Figure 2 shows the actual three-dimensional relation between the viewer, the rotating rod, and the axis about which the rod is spinning. Note that the axis of rotation (which defines the surface plane) can be any stationary vector and need not be vertical nor parallel to the xy image plane. The problem is to recover the correct axis of rotation (as well as the length of the rod).

4.2 Rigid Rod and Rotation in a Plane (\hat{P})

Let the coordinate system be centered at the projection of the midpoint of the rod. Then since the distance $OA = OA'$, we need consider the motion of OA only. Let the three-dimensional coordinates of end A be (x_1, y_1, z_1) for frame 1 and (x_i, y_i, z_i) for frame i . Then since the "stick" is a rigid rod, we have the constraint that the rod length remains constant for any frame:

$$x_1^2 + y_1^2 + z_1^2 = x_i^2 + y_i^2 + z_i^2 \quad (13)$$

For N frames, the relation (13) will yield $(N - 1)$ equations, each in two unknowns, z_1 and z_i (since x_i, y_i are observables in the image plane). So far we thus have $(N - 1)$ equations in N unknowns.

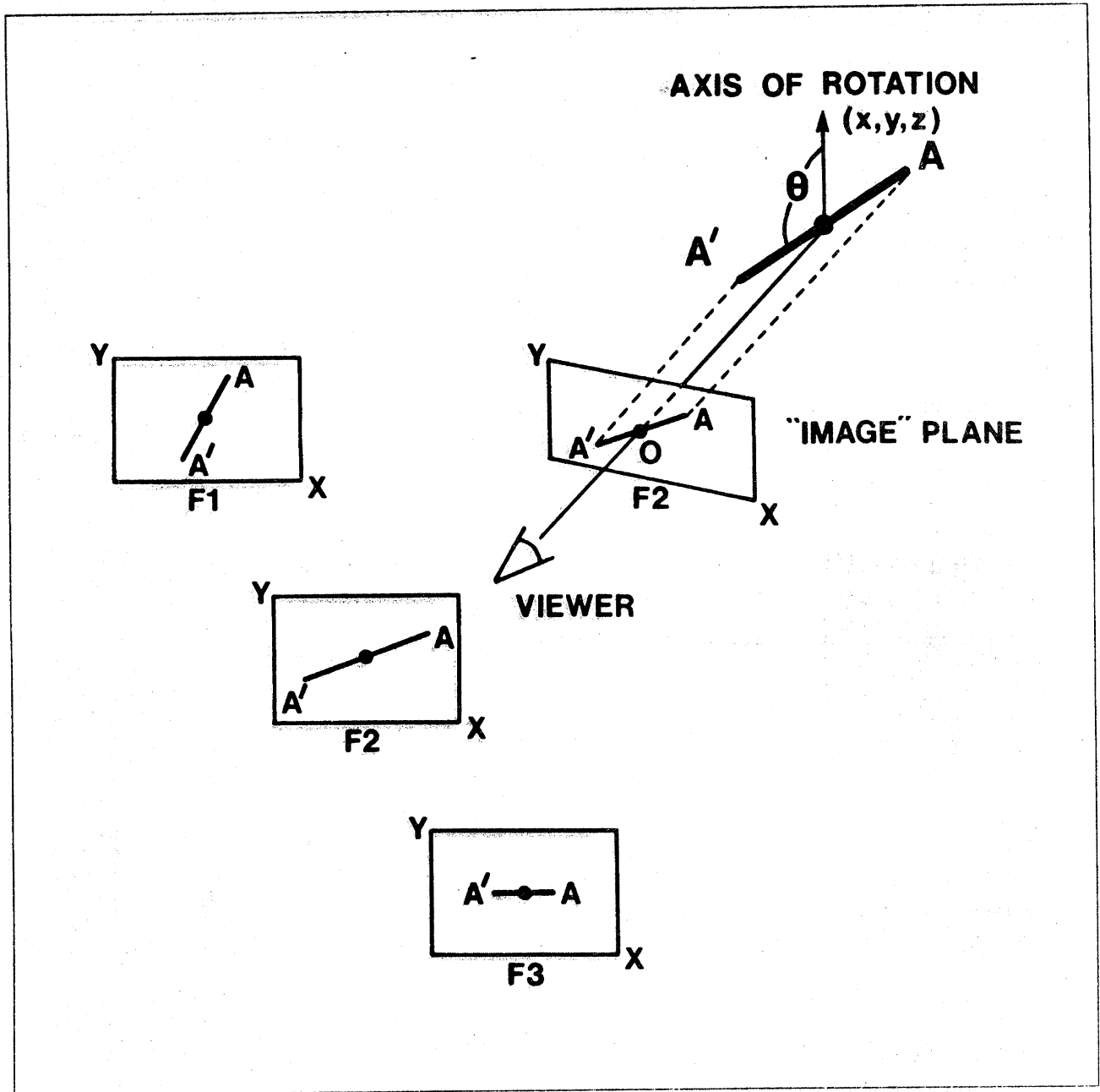


Figure 2. A simple rod rotating in three space about its midpoint.

To embody the condition of rotation about a fixed axis, we note that the angle θ between OA and its axis must remain constant. This can be expressed by forming the dot product between the rod segment OA with the presumed axis of rotation, N :

$$OA_i \cdot N = \cos \theta \tag{14a}$$

where the subscripted OA_i indicates the 2-D projection of the 3-D length OA onto the i -th frame.

Letting the end position of the unit axial vector \mathbf{N} have the coordinates x_0, y_0, z_0 , equation (14a) reduces to

$$x_i \cdot x_0 + y_i \cdot y_0 + z_i \cdot z_0 = k \cos \theta \quad (14b)$$

where $k = (x_i^2 + y_i^2 + z_i^2)^{1/2}$.

But rotation in a plane requires that the angle θ between the axis \mathbf{N} and OA be $\pi/2$. Hence, $\cos \theta = 0$ and the value of k is irrelevant. For N frames, relation (14b) thus gives us N equations in three more unknowns: x_0, y_0, z_0 . However, because the length of the rotation axis is irrelevant also, \mathbf{N} can be taken as the unit vector and we obtain the additional equation

$$x_0^2 + y_0^2 + z_0^2 = 1 \quad (14c)$$

Altogether, we thus have $(N - 1) + N + 1$ equations (E) in $N + 3$ unknowns (U): z_i, x_0, y_0, z_0 . (Note that all of these equations are polynomials.) The minimum number of equations can then be determined from the relation $E \geq U$:

$$2N \geq N + 3 \quad (15)$$

or

$$N \geq 3$$

4.3 The Jacobian Test

The next step is to demonstrate that the equations (13) and (14) form a set of independent equations. We thus examine the Jacobian for $N = 3$ to see if its rank is maintained. Recalling that x_i, y_i for $i \neq 0$ are given in the image plane, the partial derivatives of z_i in equation (13) for $i = 2, 3$ yield the first two rows of the following matrix, while the remaining rows come from from equations (14b) and (14c) respectively:

$$\begin{vmatrix} 2z_1 & -2z_2 & 0 & 0 & 0 & 0 \\ 2z_1 & 0 & -2z_3 & 0 & 0 & 0 \\ z_0 & 0 & 0 & x_1 & y_1 & z_1 \\ 0 & z_0 & 0 & x_2 & y_2 & z_2 \\ 0 & 0 & z_0 & x_3 & y_3 & z_3 \\ 0 & 0 & 0 & 2x_0 & 2y_0 & 2z_0 \end{vmatrix}$$

Evaluation of the determinant by MACSYMA shows that it is generally non-zero. However, certain relations between the variables may cause the Jacobian to drop rank. Some of these failure conditions can be noted by factoring the determinant. (Note that such failure conditions provide instances where

any perceptual system that interprets data in accord with the system of equations should also fail. The factors thus provide example experiments for "instant psychophysics".)

4.4 Bezout's Theorem and Uniqueness

Although the set of equations (13) and (14) are shown to be "independent" by the Jacobian test, Bezout's Theorem tells us that we may have up to $2^6 = 64$ possible solutions. (This is the product of the degrees of the six equations). Which of these solutions do we pick?

Fortunately, it can be shown by algebraic reduction of the six equations that of these 64 possible solutions, only two have real values—and one of these is simply a "reflection" of the other about the image plane (Hoffman and Flinchbaugh, 1981).² Thus, three snapshots or "frames" showing the x, y positions of the end points of a rotating rod are sufficient to solve for the rod length and its axis. (The reflection causes an ambiguity only in the direction of motion and orientation of the rod.) But since *any* triplet of x, y positions will yield a solution, how do we know that the measurements were taken from a rotating rod and not from a random set of points? Clearly additional tests must be performed before any meaningful interpretation can be given to the data.

4.5 Corroboration

In addition to the problem of isolating a unique solution point, it is also necessary to show that the "unique" solution is indeed plausible. (If the unique solution is not physically realizable, it can be rejected immediately.) In the case of the rod rotating in a plane about a fixed axis, three frames (or snapshots) were sufficient to solve the six polynomial equations and to obtain a unique solution for the rod's length and its axis of rotation. However, are we guaranteed that no other set of conditions could generate the data? Clearly not, for if the simple rod rotation is simulated in the laboratory on a TV monitor, then one obvious interpretation is that there are two points moving on the face of the TV. (In fact, if reflections appear on the screen so that strong 3-D cues are present, then the illusion of a rod rotating in 3-D is lost.)

Before a final interpretation should be made, it is therefore prudent to corroborate the solution to increase the probability for a correct interpretation. This can be accomplished by analyzing an independent set of data or hypotheses that are based on entirely distinct physical constraints. (In the case of structure from motion, stereopsis may be used.) Without such corroboration, the human observer seems to accept the interpretation that is most favored by the real-world statistics.³

²In the event that algebraic reduction is not possible, then the uniqueness of a solution can be tested by generating data from several known, but arbitrary configurations, and by numerical evaluation determine if the correct solution is obtained (Ullman, personal communication). Numerical evaluation is recommended in any case as a further check for the isolation of solution points.

³In the rotating rod case where the screen or reflections are not visible, then because there is no contrary 3-D information, the 3-D interpretation will be accepted as most likely.

5. Hidden Dependencies

Quite often when the equation-counting method is used, the constraint equations contain hidden dependencies that cause the Jacobian to drop rank and its determinant to equal zero. There are two general procedures for handling this situation so that an interpretation of the data can be made. The first is simply to introduce another independent constraint, the second is to identify the dependency and to reduce the number of physical variables accordingly. The disambiguation of shadows and highlights illustrates these two methods.

5.1 Example 2: Interpreting Shadows and Highlights

Consider the very common situation in vision when two patches of surface A and B appear superficially different. Do A and B differ because they have different reflectances, or is one of the regions a highlight or a shadow on a surface of uniform reflectance? These two interpretations are different, since when B is a shadowed region, the implication is that there is an object occluding the direct light of the source, whereas in the highlight case, the difference between A and B is due to the specular properties of the surface and there is no cast shadow.⁴ (If the darker region around the highlight were to be regarded as shadowed, then 99 per cent of the world would be interpreted as lying in shade!)

As shown in Figure 3, let the observer view the surface from above, and let the surface be illuminated with at least two sources of illumination—one producing direct light, as from a sun, while the other source is diffuse, such as that characteristic of the sky and clouds.

We proceed by noting that the only information available to the viewer is the image intensities I_A, I_B from the two regions A and B . For simple Lambertian conditions, these image intensities will be the product of the strength of illumination times the reflectances of the surface material. Let the reflectance common to A and B be R_λ where the subscript λ indicates R is a function of wavelength, and let S_λ be the incident flux from the direct light of the sun and D_λ the flux arising from the diffuse light from the sky, both of which are also functions of wavelength as indicated by the subscript.⁵ If a region is neither highlighted nor shadowed, then the image intensity I will be given by

$$I = (S_\lambda + D_\lambda)R_\lambda \quad (17a)$$

Equation (17a) thus describes the image intensity resulting from an unshadowed, matte surface.

5.2 The Highlight Case

⁴Note that for this analysis we are ignoring other distinctive features of a highlight: 1) the textural aspect of specularity, 2) its directional component which produces a disparity between the two eyes, and 3) that highlight edges are convex whereas shadow edges tend to be straight or concave.

⁵A planar surface is assumed; the effect of surface orientation on the source illumination can be considered incorporated into S_λ and D_λ .

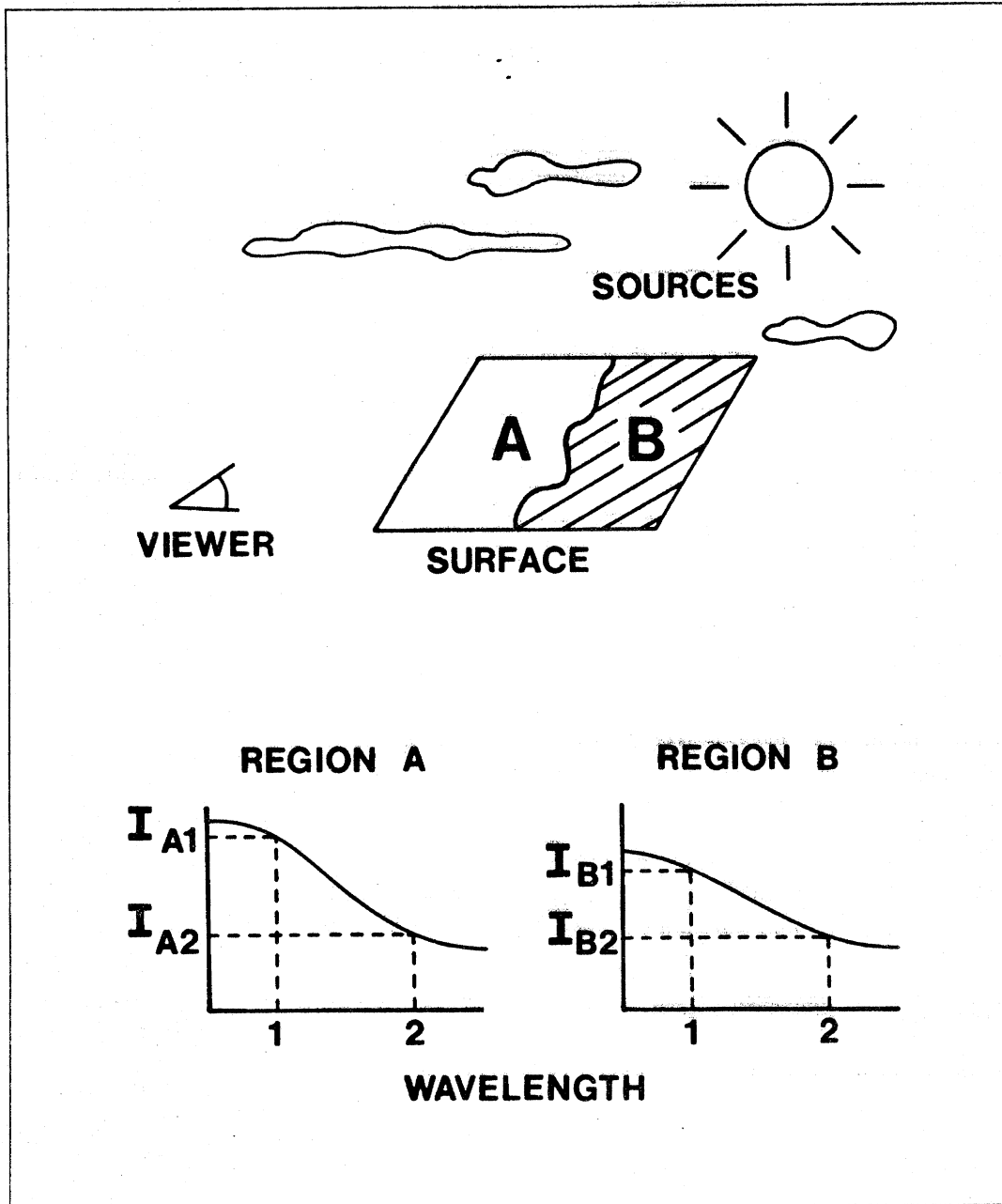


Figure 3. Direct and diffuse light illuminate the surface. Is region A a highlight or is region B in shadow? Possible image intensities over wavelength are illustrated in the lower pair of graphs.

If region A is the same flat surface as region B, except that it has a highlight, then B remains matte and I_B is defined by equation (17a). On the other hand, equation (17a) will not apply to the highlighted region A, which acts like a partial mirror reflecting some fraction of the illuminated scene lying away from the viewer. The reflectance R_λ will thus depend in part upon what the viewer sees in the reflection off A. In the case of the normal highlight, the arrangement between the direct source illumination S_λ , the surface, and the viewer is such that only the source light is reflected off the viewed surface and hence $R_\lambda = 1$ and $D_\lambda = 0$ (for the highlight only). This contribution from the

reflected light to the image intensity I_A is at the expense of the matte component of surface reflectance (Evans, 1948; Horn, 1977). Thus, if the highlight contribution to the image intensity is the fraction f_H , then the matte contribution will be $(1 - f_H)$. To characterize the image intensity I_A corresponding to a partial highlight on region A , we may thus reduce the matte equation (17a) by the factor $(1 - f_H)$ and add to it the complementary fraction f_H of specular light:

$$I_{A\lambda} = f_H S_\lambda + (1 - f_H)(S_\lambda + D_\lambda)R_\lambda \quad (17b)$$

where the first term on the R.H.S. is the specular component and the second term is the matte component of the highlight. Note that only the illuminant S_λ appears in the specular term because of the directional properties of the reflections off a highlighted region.⁶

5.3 The Shadow Case

If region B is the same surface as A , but B is in shadow, then region B will be illuminated only by the diffuse light D_λ . The effect of shadowing is thus to reduce the illumination from $(S_\lambda + D_\lambda)$ to D_λ . Recognizing that shadows often have penumbræ, we may let f_S be the fraction of the total illumination that contributes to the shaded region. For shadow, therefore, equation (17a) may be modified as follows:

$$I_{B\lambda} = f_S(S_\lambda + D_\lambda)R_\lambda + (1 - f_S)D_\lambda R_\lambda \quad (18a)$$

which further simplifies to

$$I_{B\lambda} = (f_S S_\lambda + D_\lambda)R_\lambda \quad (18b)$$

For complete shade, $f_S = 0$ and the image intensity $I_{B\lambda}$ arising from region B is described only by the product of the diffuse light times the reflectance. For no shade, $f_S = 1$; and for the penumbræ, f_S lies between 0 and 1.

5.4 Preliminary Equation Counting

Equations (17b) and (18a) may be combined to obtain a single equation that describes the image intensity for both the highlight and shadow conditions. This can be accomplished quite easily by replacing the matte component in the highlight equation (17b) by the shadow relation of (18a). After simplification, the resulting single equation will be

$$I_\lambda = f_H S_\lambda + (1 - f_H)(f_S S_\lambda + D_\lambda)R_\lambda \quad (19)$$

⁶Note that the equation describing the highlight condition is similar to that used for transparency.

where the λ subscript indicates a wavelength dependency, and f_H and f_S are respectively the highlight and shadow fractions.

If I_λ , f_H , f_S are now indexed to indicate the spatial region, we can apply the standard equation-counting procedure to determine the minimum number of wavelength and spatial samples needed to solve for the physical variables S_λ , R_λ , D_λ , f_{iH} and f_{iS} in terms of the known $I_{i\lambda}$, and then attempt to determine whether the solution for these physical variables implies a shadow or highlight.

Unfortunately, the equation-counting procedure is unsatisfactory in this case for two reasons. First, the minimum number of spatial and spectral samples is biologically unfeasible (5 and 5 or 6 and 4, respectively); second, and more important, the Jacobian collapses. The collapse is due to hidden dependencies in the set of equations of the form (19).

5.5 Eliminating Dependencies

The most obvious strategy for eliminating dependencies among equations is to search for other independent relations or constraints. Often, this may be difficult, and a more desirable course is to try to reduce the number of unknowns by combining some of the physical variables whose solution is not critical to the interpretation. For example, if the pairs $S_\lambda R_\lambda$ and $D_\lambda R_\lambda$ occur together everywhere, then we might consider replacing each pair by a single variable. Such a reduction would not affect the ability to distinguish a shadow from a highlight. Each of these two procedures will now be illustrated.

5.6 Solving for the Highlights by Adding Constraints

To introduce additional independent constraining relations, we will consider the two-dimensional case as shown in Figure 4 where a highlight (or shadow) runs across a change in reflectance R_1, R_2 . The highlight boundary is parallel to the Y axis; the reflectance change is parallel to the X axis. For this two-dimensional case, equation (19) will assume the following form:

$$I_{XY\lambda} = f_X L_{Y\lambda} + (1 - f_X) M_{Y\lambda} \quad (20)$$

where $I_{XY\lambda}$ is the image intensity corresponding to one of the regions A_1, B_1, C_1 or A_2, B_2, C_2 . Note that since only two wavelength variables L_λ and M_λ are involved along the X axis, these variables need to be indexed by Y only.

By simple equation-counting, it can be verified that the minimum number of samples along X or Y and for λ will be respectively either 3,1,3 or 3,3,1. (Note that Y and λ appear together and hence can be symmetrically indexed). A further reduction can be obtained by noting that region C_1 or C_2 , etc. is always matte, and hence f_C (or $f_{x=3}$) is zero. Thus $I_{C\cdot Y\lambda} = M_{Y\lambda}$. The minimum for X, Y, λ is then 3, 1, 2 or 3, 2, 1, which correspond to a set of six equations in six unknowns. The determinant of the Jacobian of either system of equations is still zero, however.

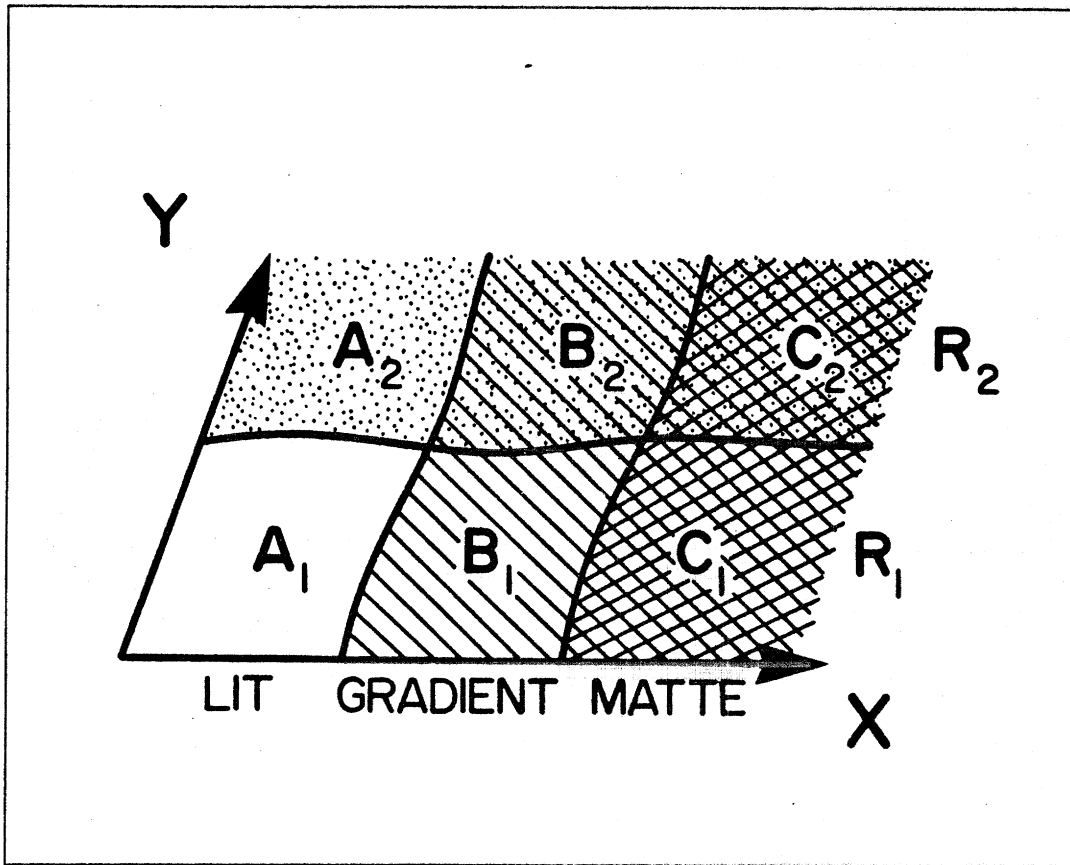


Figure 4. View of a surface with a shadow or highlight boundary parallel to the Y axis and crossing a region of two different reflectances, R_1 and R_2 .

To solve the equations, we need to introduce one more constraint or reduce the number of variables. For highlights, an additional constraint can be added by noting that the spectral composition of the purely specular component is independent of the underlying reflectance R_1, R_2 . Thus along Y , $L_{i\lambda} = L_{j\lambda}$. The minimum X, Y, λ samples are now $X = 2, \lambda = 1, Y = 2$ (the symmetry between Y and λ has been removed by the specular constraint), leading to the following equations:

$$I_{B1} = f_B L_1 + (1 - f_B) M_1$$

$$I_{B2} = f_B L_2 + (1 - f_B) M_2$$

$$I_{C1} = f_C L_1 + (1 - f_C) M_1 \quad (21)$$

$$I_{C2} = f_C L_2 + (1 - f_C) M_2$$

$$f_C = 0 \quad L_1 = L_2$$

where the indexing is for Y only, since there is only a single wavelength sample.

The Jacobian of the reduced set of the above equations obtained by substituting $L_1 = L_2$ and $f_c = 0$ is:

$$\begin{vmatrix} L_1 - M_1 & f_B & (1 - f_B) & 0 \\ L_1 - M_2 & f_B & 0 & (1 - f_B) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = f_B(M_2 - M_1)$$

which is non-singular provided $M_1 \neq M_2$ and $f_B \neq 0$.

Thus solutions can be obtained for f_B , M_1 , M_2 and particularly L_1 , the specular component of the light reflected off the surface.

$$L_{\text{specular}} = \frac{I_{B2}I_{C1} - I_{B1}I_{C2}}{(I_{C1} - I_{C2}) - (I_{B1} - I_{B2})} \quad (22a)$$

$$1 - f_B = \frac{I_{B1} - I_{B2}}{I_{C1} - I_{C2}} \quad (22b)$$

5.7 Solving for Shadows by Combining Variables

Returning to Figure 4, we may now reinterpret the regions A_1 , A_2 , B_1 , B_2 in terms of a shadow edge parallel to the Y -axis. (A penumbra will be needed for this constraint implying that the minimum spatial samples along X is three although only two will be used as in the highlight case.)

For shadows, the equation (19) then has the same form as the first four equations (21), with $L_i = (S_i + D_i)R_i$ and $M_i = D_iR_i$, where S and D are respectively the source and diffuse light and R is the reflectance. Since for shadows $L_1 \neq L_2$ (i.e., there is no spectral component superimposed on A_1 , A_2 , or B_1 , B_2), an additional constraining equation must replace this specular constraint. For illustration, we will introduce a "gray world" condition, namely that the average of all surfaces reflecting the source light is spectrally flat. Hence the diffuse light D_i is simply some fraction γ of the source light:

$$D_i = \gamma S_i \quad (23a)$$

and

$$M_i = \gamma S_i R_i \quad (23b)$$

$$L_i = (1 + \gamma) S_i R_i \quad (23c)$$

Because S and R appear together in two of the above equations, they cannot be solved for separately, and the Jacobian test will fail when applied to equations (23). To eliminate this dependency, define a new variable $S^* = S \cdot R$. The shadow equations then become:

$$I_{B1} = f_B(1 + \gamma)S_1^* + (1 - f_B)\gamma S_1^* = (f_B + \gamma)S_1^*$$

$$I_{B2} = f_B(1 + \gamma)S_2^* + (1 - f_B)\gamma S_2^* = (f_B + \gamma)S_2^*$$

$$I_{C1} = \gamma S_1^* (= M_1)$$

$$I_{C2} = \gamma S_2^* (= M_2)$$

(24)

with the four unknowns being f_B , γ , S_1^* , S_2^* .

Unfortunately, the determinant of the Jacobian of this set of equations is still zero, suggesting that dependencies are still present:

$$\begin{vmatrix} (f_B + \gamma) & 0 & S_1^* & S_1^* \\ 0 & (f_B + \gamma) & S_2^* & S_2^* \\ \gamma & 0 & 0 & S_1^* \\ 0 & \gamma & 0 & S_2^* \end{vmatrix} = 0$$

Rather than introducing a new constraint, we will proceed to determine whether any of the physical variables can be combined to reduce further the number of unknowns. The most obvious choices are ratios or products of the entries in the Jacobian array. These terms are the coefficients of the variables in the original set of equations, and consequently are the factors that would be used to multiply two of the equations to eliminate one variable. (In essence, we are exploring various triangular forms of the matrix of rank one less than the original.) The appropriate ratios are thus those between the rows in the same columns, because it is these factors that will be cross multiplied to eliminate the variable that is identified with that column of the Jacobian matrix. Thus the appropriate ratios of the above Jacobian that should be explored first are $(f_B + \gamma)/\gamma$, which appear in columns 1 and 2, and S_1^*/S_2^* , which appear in columns 3 and 4. Inspection of equations (24) shows that the solution for these reduced variables is quite simple

$$\frac{S_1^*}{S_2^*} = \frac{I_{B1}}{I_{B2}} = \frac{I_{C1}}{I_{C2}} = \frac{S_1 R_1}{S_2 R_2} \quad (26a)$$

$$\frac{f_B + \gamma}{\gamma} = \frac{I_{B1}}{I_{C1}} = \frac{I_{B2}}{I_{C2}} \quad (26b)$$

The extra solution for each paired variable now reveals the dependency between the image intensities that caused the rank reduction of the Jacobian of (24), namely the relation

$$I_{B1}I_{C2} = I_{C1}I_{B2} \quad (26c)$$

which is common to both (26a) and (26b). If the grey world condition applies and if $C(Y)$ is a shadow on $B(X)$, then the shadow relation (26c) will be true.

Unfortunately, there are an unlimited number of image intensity values that will satisfy the "shadow" relation (26c). How are we to be sure that they all correspond to the shadow condition and not to a reflectance change or even a highlight? To answer this question, we proceed in two stages, first to show that the shadow solution (26) never will correspond to a highlight, and hence shadows and highlights are at least disambiguated because their solutions are distinct. Then, we will illustrate how the probability of other confounding spectral relations such as different materials can be set arbitrarily low by independent corroboration of the original solution.

6. Distinctness of S and H Solutions (Exclusion of Competing Interpretations)

Our basic procedure to prove distinctness of the shadow S and highlight H solutions will be to show that there is at least one relation between the four available image intensities (I_{B1} , I_{B2} , I_{C1} , I_{C2}) that has different values for the shadow and highlight conditions. These values will always be different (if the constraints are valid) because the relation corresponds to two different physical variables (one for shadow, the other for highlights) that have non-overlapping values.

To proceed, we ask first what highlight conditions satisfy the shadow solution (26). (Subsequently, we will examine the opposite case—asking what shadow conditions will "look like" highlights.)⁷ We thus assume relation (26) holds and solve for one of the highlight conditions. Consider equation (22a) that specifies the magnitude of the specular components of the highlight. Note the numerator is identical to the shadow equation (26) if the left hand side (L.H.S.) of (26) is subtracted from the R.H.S. In this case, however, the numerator (22a) will be zero. Hence the shadow condition requires that $L_{\text{specular}} = 0$ and consequently there can be no highlight interpretation. Thus, given that the shadow condition (26) holds, there will be no highlight interpretation.

To check for the reverse case, namely under what conditions the image intensity relations for the highlight condition will also yield a shadow interpretation, we may examine the second highlight equation (22b). In particular, we wish to solve for the physical interpretation of the intensity relations of (22b) given a shadow condition. This can be accomplished simply by substituting equations (24) into the R.H.S. of (22b). We find that, given the shadow conditions, then

$$\frac{I_{B1} - I_{B2}}{I_{C1} - I_{C2}} = \frac{f_B + \gamma}{\gamma} = \frac{f_B}{\gamma} + 1 \quad (27)$$

Figure 5 now plots the possible values of the image intensity ratio given by the LHS of (27) for shadows and the RHS of (22b) for highlights.

⁷For another example treatment, see Ullman's (1979) analysis of false-targets for his structure-from-motion theorems.

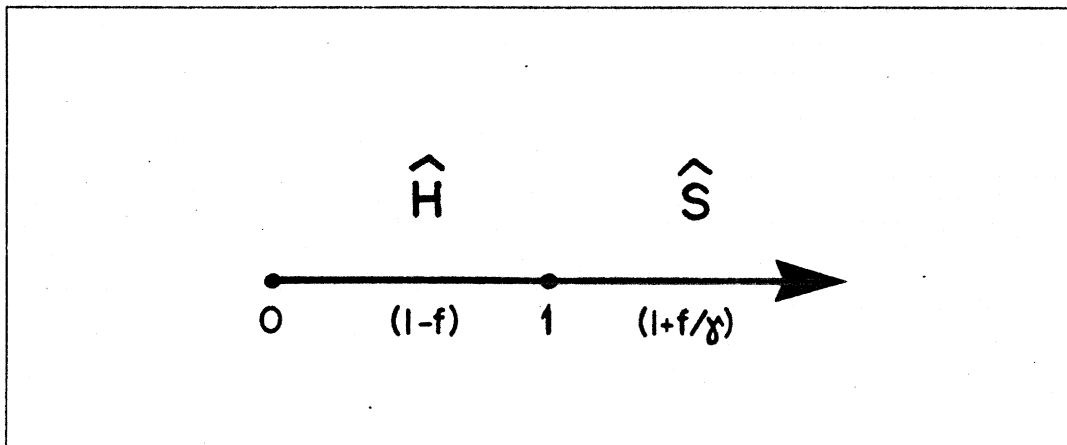


Figure 5. Solution space for shadow S and highlight H conditions.

We note that both f (the fraction of specularity or shadow) and γ (the fraction of direct light), range between 0 and 1. Hence for highlights, $1 - f$ must lie between 0 and 1, whereas for shadows $1 + f/\gamma$ will be greater than or equal to 1. The only common condition is when $f = 0$, which corresponds to a homogeneous matte area. Thus, highlights and shadows will never be confused from the image intensities (provided the gray world assumption applies), if the calculation given by the L.H.S. of (27) is made. It is of some interest that this operation on image intensities is equivalent to examining the output of the double-opponent color cell found in most biological color vision systems (see Rubin and Richards, 1981).

6.1 Corroboration

Although the highlight H and shadow S solutions are unique and distinct, it is still possible that other properties of surfaces, such as pigment density changes or changes in reflectances could satisfy equations (22) or (25) and be misinterpreted as either a highlight H , or shadow S . Thus a shadow or highlight interpretation should not yet be given to the solutions H and S . To exclude *all* other possibilities is difficult (see Rubin and Richards, 1981, however). Nevertheless, the odds for an incorrect H or S interpretation can be reduced by applying an independent test for the validity of the shadow or highlight equations. We call such a procedure "corroboration".

One simple independent corroborative test is to note whether the equation counting procedure suggested more than one minimal condition for solution. In particular, we noted in section 5.5 that the equation (20) had a symmetry in wavelength (λ) and space (Y). We chose as a starting point one spectral sample and two samples in the Y dimension. An independent test would therefore be to use two spectral samples rather than one, and only one sample in the Y dimension. This case corresponds to examining the gradients of a highlight, or the penumbra of a shadow.

A second and more common type of corroborating procedure is to simply take another set of measurements independent of the first, and determine whether the solutions for the physical constants remain the same or not. If they do not, then the interpretation must be rejected. If they are confirmed,

then the odds on a misinterpretation are reduced. Ideally, the corroboration should be based upon measurements taken from a physical dimension different from that used in the original solution. In any case, since we are corroborating the value of a physical parameter, the corroborating measurements must not be confounded with the dimensions of that physical parameter. In this respect, the relation (27) that tests for the highlight or shadow condition is most satisfactory, for the values f_B and γ are dimensionless and are not functions of wavelength, for example. For the shadow condition, we thus can take a third spectral sample I_{B3}, I_{C3} and substitute these image intensities for I_{B2}, I_{C2} . Since the physical constant $(f_B + \gamma)/\gamma$ of equation (25b) is not a function of wavelength, this value should remain unchanged if the image intensity changes are indeed due to a shadow. In effect, we are confirming that the S solution point remains fixed along the solution ray illustrated in Fig. 5. If it does, then the shadow (or highlight) interpretation is reaffirmed and the chance of misinterpretation is unlikely provided that the competing interpretations are not processes that behave like shadows. Consequently, at least three wavelength samples are required before a reliable shadow interpretation can be made.

In the case of recovering structure from motion—our earlier example—the corroboration of the axis of rotation could entail adding additional *frames* or snapshots to see if the same axis and rod length is recovered. Clearly, this procedure is not entirely independent because the strategy for solution remains the same and some possible confounding interpretations may not be excluded (e.g., the correct interpretation that the points are on a TV monitor in 2-D).

A more independent corroborative test would be to use stereopsis, for this computation of the depth relations between the feature points is quite different from the structure-from-motion analysis. This ideal corroborative procedure should thus use an entirely different computational analysis, which is based upon relations that have quite different failure conditions.⁸

7. Summary

Although the equation-counting procedure has been used in the past to give some insight into the complexity required to solve problems in many non-linear variables (e.g., Leith *et al*, 1981), researchers in perception have often neglected to recognize that certain other conditions must be fulfilled before a meaningful solution can be guaranteed (Meiri, 1980). These conditions are summarized in the flow diagram of Fig. 6. They include the Jacobian test for the independence of the system of equations, uniqueness of solution, exclusions of competing interpretations, and two kinds of corroboration. If these conditions can be met, then the equation counting procedure provides a powerful theoretical tool for understanding how, in principle, biological systems can make reliable interpretations and assertions from the greatly impoverished sense data available to them.

⁸For biological systems, we probably should view "corroboration" as an early step in the perceptual process (perhaps at the level of Marr's 2-1/2D sketch) that acts on the output of modules analyzing information derived from motion, disparity, color, texture, etc., as well as non-visual information, such as tactile roughness, shape or even in some cases, acoustic information.

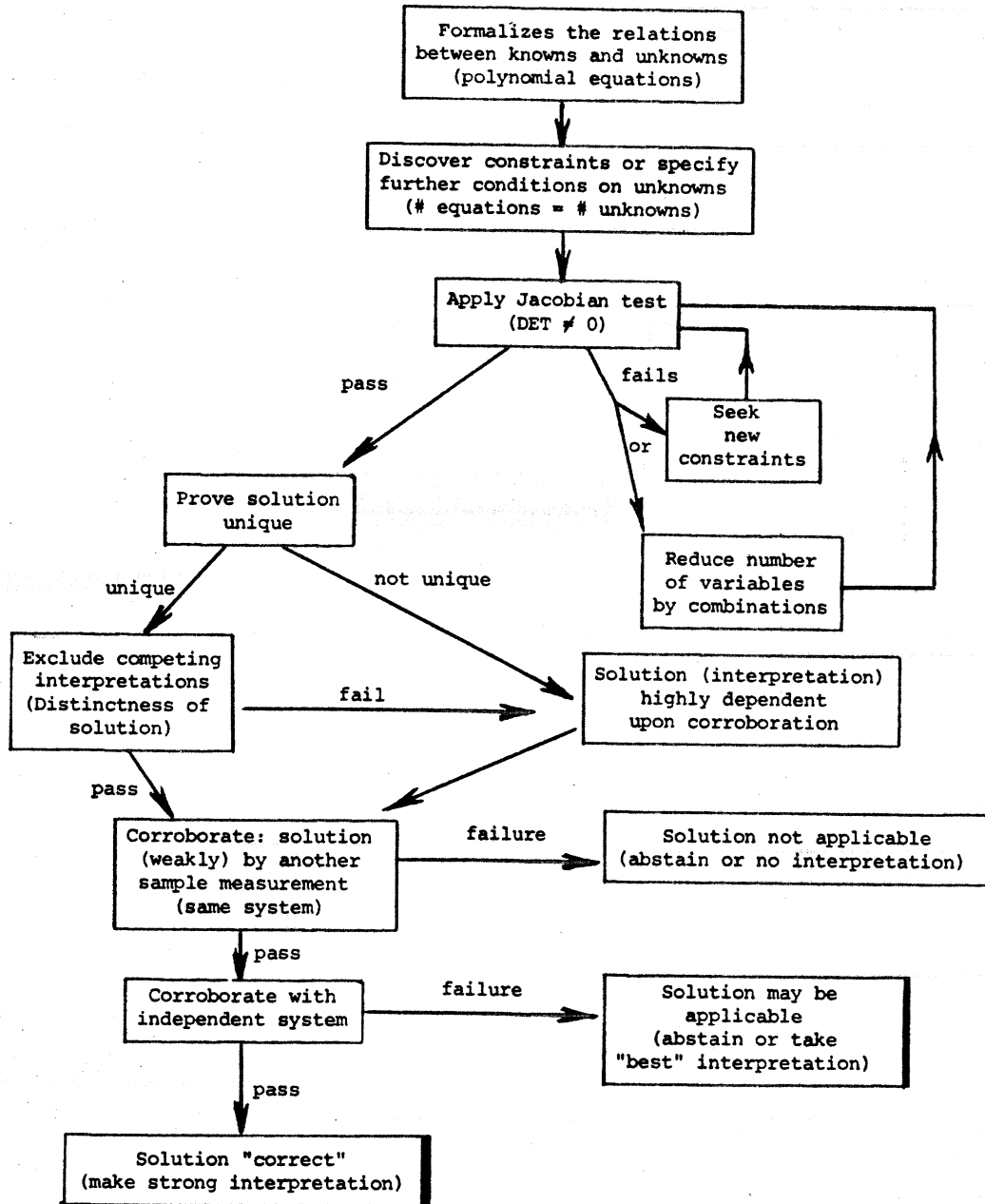


Figure 6. Outline of Steps in 'Equation-Counting' Procedure.

Appendix I: Redundancy

Unfortunately, due to measurement and sampling errors, real-world data are not precise. The hardware performing the calculations may also be quite noisy, as is the case for many neural networks. Without exact data and calculations, solution vectors will not be completely isolated, but rather are more properly represented as a probability distribution about the exact solution point. To reduce the likelihood of misinterpretation, several overconstraining equations are often helpful. (By "overconstraining" we here mean the inclusion of equations in addition to those needed to obtain a unique solution.) Their value will depend in part upon how many variables (unknowns) are included in the solution point. Intuitively, the more the unknowns, the greater the potential noise and the less the contribution of any one overconstraining equation will be. To capture this property, we suggest the following measure of the redundancy of a system containing overconstraining equations:

$$\text{Redundancy} = 1 - \left[1 - \frac{1}{U}\right]^C \quad (\text{A1})$$

where C is the number of independent combinations of the equations and U is the number of unknowns. As U increases, this measure decreases to zero. The effect of the additional overconstraining equations, on the other hand, is to reduce the deleterious effect of increasing U in a manner analogous to probability summation, yet the redundancy measure will never exceed 1 (the ideal). The redundancy measure has the practical value of providing an estimate of how many extra equations (or data samples) are needed to isolate a solution point to a certain probability, given known measurement signal to noise ratios.

Appendix II: Sard's Theorem for non-Polynomial Functions

In many cases, the equations relating the unknown variables will not be polynomial and Bezout's Theorem will not apply. These exceptions include such common functions as exponentials, logarithmic, or trigonometric. Sometimes, a change of variables can be made to recast the non-polynomial relations in polynomial form. If this is done, then care must be taken to restrict the range over which the polynomial form applies.

More generally, if a function is smooth on a manifold, then Sard's Theorem can be used (Guillemin and Pollack, 1974; Milnor, 1978). Suppose that the following system of independent equations holds:

$$f_1(x_1, \dots, x_k) = p_1$$

$$f_n(x_1, \dots, x_k) = p_n$$

This system can then be represented more generally as a mapping from R^k to R^n :

$$F: R^k \mapsto R^n$$

or

$$F(x_1, \dots, x_k) = \{f_1(x_1 \dots x_k), \dots, f_n(x_1 \dots x_k)\}$$

By Sard's Theorem, we know that if F is a smooth mapping and if F is invertible for the values p , then the dimension of $F^{-1}(p)$ is $(k - n)$. Since when $k = n$ the dimension of $F^{-1}(p)$ is zero, there can be at most a countable number of (isolated) solutions.

Some care must be taken in assuming that Sard's Theorem applies to any differentiable function. It does not. For example, consider the simple periodic function $\sin x$. Such a function is uniquely invertible only over a specified range. Polynomial functions are thus a "safer" class of functions to use for equation counting, for their appropriate range is usually more obvious.

REFERENCES

- Clowes, M., "On Seeing Things," *Artif. Intell.* 2 (1971), 79-116.
- Evans, R.M., *An Introduction to Color*, J. Wiley, New York, 1948.
- Guillemin, V. & Pollack, A., *Differential Topology*, Prentice-Hall, Englewood Cliffs, New Jersey, 1974.
- Hoffman, D. D. & Flinchbaugh, B. E., "Interpreting Biological Motion," *Biological Cybernetics*, in press (also MIT AI Memo 608) (1981).
- Horn, B., "Understanding Image Intensities," *Art. Intell.* 8 (1977), 201-231.
- Huffman, D., "Impossible objects as nonsense sentences", *Machine Intelligence* 6, B. Meltzer & D. Michie, eds., Edinburgh University Press, 1971.
- Johansson, G., "Visual Motion Perception," *Scientific American* (1975), 76-88.
- Kendig, K., *Elementary Algebraic Geometry*, Univ. Press of Virginia, Charlottesville, 1977.
- Leith, E. N., Chen, H., and Cheng, Y. S., "Diffraction-limited imaging through a phase-distorting medium," *Optics Letters* 6 (1981), 4-6.
- Marr, D. C., *Vision: A computational investigation into the human representation and processing of visual information.*, W. H. Freeman & Co., San Francisco, 1982.

Marr, D. C., "Early processing of visual information," *Phil. Trans. R. Soc. Lond. B.* 275 (1976), 483-524.

Meiri, A. Z., "On monocular perception of 3-D moving objects," *IEEE Trans. on Pattern Analyses & Machine Intell.* PAMI-2 (1980), 582-583.

Milnor, J. W., *Topology from Differentiable Viewpoint*, Univ. Press of Virginia, Charlottesville, 1978.

Pentland, A., "Finding the illuminant direction," *MIT AI Memo No. 584* (1980).

Rubin, J. M. & Richards, W. A., "Color vision: a computational view," *A. I. Memo 631* (1981).

Thomas, G. B., *Calculus and Analytic Geometry*, Addison Wesley, Mass., 1951.

Ullman, S., *The Interpretation of Visual Motion*, MIT Press, Cambridge, Mass., 1979.

Van der Waerden, B. L., *Modern Algebra, Vol. II*, F. Ungar Publ. Co., New York, 1940.

Waltz, D., "Understanding line drawings of scenes with shadows", *The Psychology of Computer Vision*, P. Winston, ed., McGraw Hill, New York, 1975.

Witkin, A., "Shape from Contour," *MIT AI Technical Report 589* (1980).