MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ARTIFICIAL INTELLIGENCE LABORATORY

AIM 335 August 1975

# IMAGE INTENSITY UNDERSTANDING

by

Berthold K. P. Horn

ABSTRACT. Image intensities have been processed traditionally without much regard to how they arise. Typically they are used only to segment an image into regions or to find edge-fragments. Image intensities do carry a great deal of useful information about three-dimensional aspects of objects and some initial attempts are made here to exploit this. An understanding of how images are formed and what determines the amount of light reflected from a point on an object to the viewer is vital to such a development. The gradient-space, popularized by Huffman and Mackworth is a helpful tool in this regard.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643.

TABLE OF CONTENTS


Introduction.................................................. 1

Developing the Tools.......................................... 3

Image Formation............................................... 3

Projection.................................................... 3

Surface Orientation........................................... 5

Image Intensity............................................... 6

Single Point Source............................................ 7

Reflectivity Function......................................... 8

Tying It All Together......................................... 9

Gradient-space................................................ 9

Tri-hedral Corners............................................11

Gradient-space Image..........................................12

Mat Surfaces and Point-Source Near Viewer.....................13

Uniform Illumination..........................................14

Incident, Emittance, and Phase Angles.........................16

Specularity...................................................19

Finding p and q from I, E, and G..............................21

Metallic Surfaces.............................................23

Mutual Illumination...........................................25

Two Semi-Infinite Planes......................................25

Two Truncated Planes..........................................28

The Main Results....................................................31

The Semantics of Edge-Profiles......................................31

Mutual Illumination.................................................34

Obscuration.........................................................35

Determining the Three-dimensional Structure of Polyhedral Scenes...37

Lunar Topography....................................................39

Lunar Reflectivity Function.........................................40

Lunar Gradient-Space Image..........................................41

Finding a Surface Profile by Integration............................43

Finding the Whole Surface...........................................44

Ambiguity in Local Gradient.........................................45

Low Sun-Angles......................................................46

Dealing with Shadows................................................46

Generalization to Perspective Projection............................47

General Reflectivity Functions......................................50

Interpretation in Terms of the Gradient-Space.......................52

Generalization to Near Source and Near Viewer.......................53

Acknowledgements....................................................53

References..........................................................54

Mathematical Details................................................55

Dual-Space..........................................................55

The Gaussian Sphere.................................................56

The Gradient-Line is Perpendicular to the Image-Line................58

Integral of $\cos^n(\theta)$ over a Hemisphere.............................59

The Off-Specularity Angle....................60

Gradient-Space Image for a Metallic Surface....................62

Stereographic Projection....................63

A Useful Discriminant....................65

The Azimuth Angle....................67

Expanding the Discriminant....................69

Finding p, and q, given I, E, and G....................70

Determination of Orientation of Planes....................72

Mutual Illumination -- Planes Truncated in One Direction....................75

Integral Equation....................77

Iterative Solution....................79

Integration by Parts....................80

Numerical Solution....................81

## AND NOW FOR SOMETHING COMPLETELY DIFFERENT

A case will be made for the usefulness of image intensities or gray levels. Usually one would like to forget about image intensities as soon as possible, extracting only edge-fragments or regions before going on. Much of the work in image analysis has used image intensities only to segment the image, based on differences in average image intensity or some higher-order measure. A great deal of information is contained in the image intensities, however, and there are ways of exploiting this fact.

Our approach is based on the belief that it is important to understand the image-forming process if one is to construct models of the world being imaged. It is not sufficient to try an assortment of statistical, computational, or signal-processing tricks that come out of a bag of procedures that has proved useful in some other domain.

Using an understanding of the visual effects of edge imperfections and mutual illumination, we will be able to suggest interpretations of lines based on image intensity profiles across edges. A "sharp peak" or edge-effect will imply that the edge is convex, a "roof" or triangular profile will suggest a concave edge, while a step-transition or discontinuity accompanied by neither a sharp peak nor a roof component will most likely be an obscuring edge. This latter hypothesis is strengthened significantly if an "inverse peak" or negative edge-effect is also seen.

Next we will show that the image intensities of regions meeting at a joint corresponding to an object corner allow one to determine fairly accurately the orientation of each of the planes meeting at the corner. The three-dimensional structure of a polyhedral scene can thus be established without the use of size- or support-hypotheses or a finite catalogue of models.

Finally we will turn to curved objects and show that their shape can be determined from the intensities recorded in the image. The approach to this problem presented here is supported by geometric arguments and does not depend on methods for solving first-order non-linear partial differential equations. It is instead a synthesis of the previous shape-from-shading method and the gradient-space approach. [4,2].

The results presented here depend to a large degree on geometric insight gained by using the gradient-space approach popularized by Huffman and Mackworth. [1,2,3,9]. Approaching the image analysis problem in the way proposed in this paper leads to the ability to prove or disprove that certain features can be extracted from images. It is not claimed, however, that it makes any inroads on the scene analysis problem.

# DEVELOPING THE TOOLS

## IMAGE FORMATION:

Our usual visual world consists of opaque bodies immersed in a transparent medium. Since we cannot see into opaque objects, their surfaces are important for recognition and description purposes. This special nature of our visual world makes it reasonable to attempt to derive a model of what is being seen from an image. The dimensionalities of the two domains match: On the one hand, we have two-dimensional surfaces plus depth, on the other, two image dimensions plus intensity.

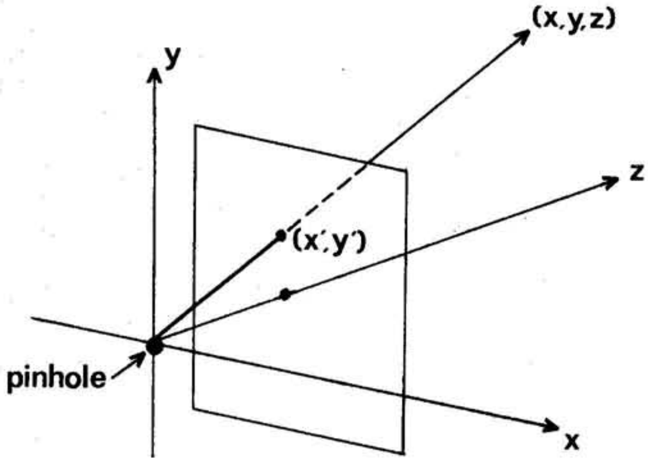
If we are to exploit this observation we have to understand how images are formed. There are two parts to this problem. One deals with the two image dimensions and relates them to the surface coordinates, and the other deals with the determination of what intensity will be recorded in the image at each point.

## PROJECTION:

First, let us look at the geometry of projection. For this purpose one can replace a lens with a pin-hole at its center. Straight lines then connect points on the objects to their images -- these lines pass through the pin-hole. If we let $(x,y,z)$ be the coordinates of some point before the viewer, and $(x',y')$ its image coordinates, then

$$x' = (x/z)f \text{ and } y' = (y/z)f$$

Here f is the separation of the image plane from the lens. It is convenient to superimpose the image plane onto the object space as follows:




Above is the well-known perspective projection. Sometimes it is convenient to consider a simpler case where objects are very far away relative to their size. We can imagine looking at them through a telephoto lens. The scene then will occupy a small visual angle and the distance to points on the object will be almost constant in the projection equation.

$$x' = (f/z_o)x \text{ and } y' = (f/z_o)y$$

This corresponds to orthographic projection.

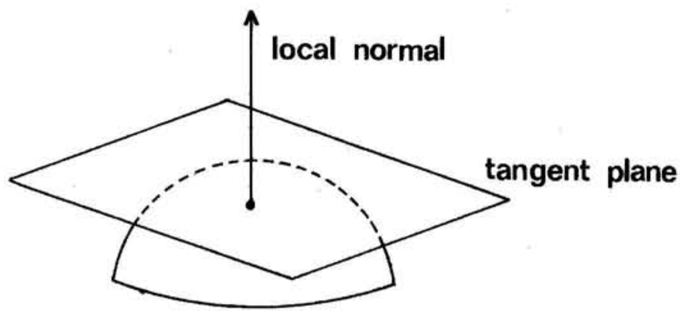
SURFACE ORIENTATION:

In order to determine the light flux reflected in the direction of the viewer from a particular surface element of the object, we will have to understand the light-source, object-surface, viewer geometry. In particular, the surface orientation will play a major role.

There are various ways of specifying the surface orientation for a plane. We can, for example, give the equation defining the plane, or the direction of a vector perpendicular to the surface. If an equation for the plane is $ax + by + cz = d$, then a suitable surface normal is $(a,b,c)$. In fact we can rewrite the equation $(x,y,z)\cdot(a,b,c) = d$. To show that any line in the surface is indeed perpendicular to the normal so defined, consider any pair of points in the surface $(x_o,y_o,z_o)$ and $(x_1,y_1,z_1)$. Connecting them and taking dot-products we find that $(x_o-x_1,y_o-y_1,z_o-z_1)\cdot(a,b,c) = 0$.

Since we shall be interested in curved surfaces as well, we extend this method for specifying surface orientation by applying it to tangent planes. That is, the orientation of the surface at a point $(x_o,y_o,z_o)$ is defined to be the orientation of the tangent plane constructed at that point. If the equation of the surface is given as $z = z(x,y)$, we can take an infinitesimal step $(dx, dy, dz)$ in the surface and find that $dz = z_x dx + z_y dy$, where $z_x$ and $z_y$ are the first partial derivatives of $z$ with respect to $x$ and $y$ respectively. Clearly, the equation of the tangent plane can be written as $z_x x + z_y y - z = d$ (where $d = z_x x_o + z_y y_o - z_o$). We can immediately construct a local normal $(z_x,z_y,-1)$.

It will be convenient to abbreviate the first partial derivatives as p and q. The local normal then becomes (p,q,-1). It is clear that orientation defined in this way has but two degrees of freedom. The quantity (p,q) will be called the gradient.




## IMAGE INTENSITY:

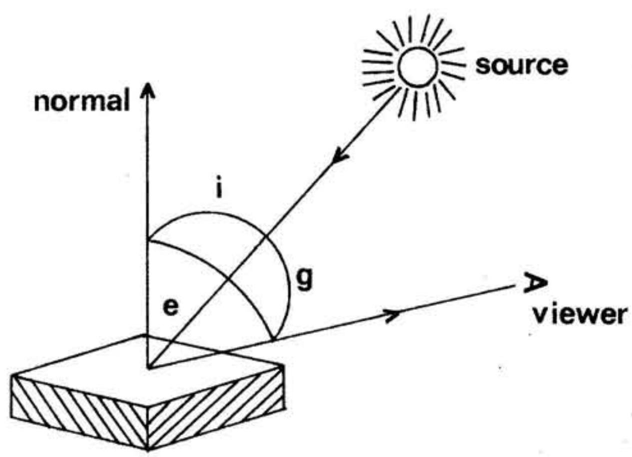
Next we turn to the image intensity. This will be equal to the amount of light reflected by the corresponding point on the object in the direction of the viewer, multiplied by some constant factor that depends on the parameters of the image-forming system. To be precise, we have to think of intensity as light flux per unit area and correspondingly also have to consider the reflected light per unit area as seen by the viewer.

Now the amount of light reflected by a surface depends on its micro-structure and the distribution of the incident light. Constructing a tangent-plane to the object's surface at the point under consideration, one sees that light may be arriving from directions distributed over a hemi-sphere. One can consider the contributions from each of these directions separately and superimpose the results.

The important point is that no matter how complex the distribution of light-sources, and for most kinds of surfaces, there is a unique value of reflectance, and image intensity, for a given orientation of the surface. We shall spend some time exploring that and develop the gradient-space image in the process.

SINGLE POINT SOURCE:

The simplest case is that of a single point-source. It is easy to see that the geometry of reflection in this case is governed by three angles, the incident, the emittance, and the phase angles. The incident angle is the angle between the incident ray and the local normal, the emittance angle is the angle between the emitted ray and the local normal, and the phase angle is the angle between the incident and emitted ray [4].




Clearly the cosines of the three angles can be found simply by taking the dot-product of the appropriate pair of unit vectors. The reflectivity function is a measure of how much of the light incident on a surface element is reflected in a particular direction. Roughly speaking, it is the fraction of the incident light reflected per unit surface area, per unit solid angle, in the direction of the viewer.

Let the illumination be E (flux/unit area) and the resulting surface luminance in the direction of the viewer be B (flux/steradian/projected area). Projected area is simply the equivalent area if the surface was not foreshortened, that is, if it was normal to the view-vector. The reflectivity is simply defined as B/E. It is usually written $\phi(i,e,g)$.

Note that an infinitesimal surface element, dA, captures a flux E cos(i)dA, since its surface normal is inclined i, relative to the incident ray. Similarly, the intensity I(flux/steradian) equals B cos(e)dA, since the projected area is foreshortened by the inclination of the surface normal relative to the emitted ray.

## REFLECTIVITY FUNCTION:

Mathematical models have been constructed for some surfaces that allow an analytical determination of the reflectivity function. Such techniques have not proved very successful so far.

In general we may not just have a single point-source illuminating the object -- other objects around it, for example, will contribute to the incident light. In this case, one has to integrate the product of the reflectivity function and the incident light per unit solid angle over the hemi-sphere visible from the point under consideration in order to determine the total light flux reflected in the direction of the viewer.

## TYING IT ALL TOGETHER:

So far we have treated geometry and intensity separately. The normal to the surface relates object geometry to image intensity. The normal is defined in terms of the surface geometry, and it also appears in the equation for the reflected light intensity since the three angles determining reflectivity depend on it. One could now proceed to develop partial differential equations based on this observation -- it is more fruitful to introduce another tool first, gradient-space. This will allow us to gain valuable intuitive insight into how one can exploit the detailed understanding of image formation.

## GRADIENT SPACE:

Gradient-space can be derived as a projection of dual-space or of the Gaussian sphere, but it is easier for our purposes here to relate it directly to surface orientation [2]. We will concern ourselves with orthographic projection only, although some of the methods can be extended to deal with perspective.

The mapping from surface orientation to gradient-space is straight-forward. If we construct a normal $(p,q,-1)$ at a point on an object, it maps into the point $(p,q)$ in gradient-space. Equivalently, one can imagine the normal placed at the origin and determine its intersection with a plane at unit distance form the origin. If we write the equation for the surface $z = z(x,y)$, then a normal to the surface will be $(z_x, z_y, -1)$, where $z_x$ and $z_y$ are the first partial derivatives of z with respect to x and y respectively. Clearly, $p = z_x$ and $q = z_y$.

We need to look at some examples to gain a feel for gradient-space. Evidently a plane maps into a point in gradient-space. A second plane parallel to the first maps into the same point. What plane maps into the point at the origin in gradient-space? A plane with normal (0,0,-1), that is, a plane perpendicular to the view-vector (0,0,-1).

Moving away from the origin in gradient-space, one finds that the distance from the origin corresponds to the inclination of the plane with respect to the view-vector -- specifically, the distance from the origin equals the tangent of the angle between the surface-normal and the view-vector, tan(e).

If we rotate the object-space about the view-vector, we induce an equal rotation of gradient-space about the origin. This allows us to line up points with the axes and so simplify analysis. Using this technique it is easy to show that the angular position of a point in gradient-space corresponds to the direction of steepest descent on the original surface.ace.

Let us call the orthogonal projection of the original space, image-space. Usually this is all that is directly accessible to us. Two planes intersect in a line. Let us call the projection of this line the image-line. The two planes, of course, also correspond to two points in gradient-space. The line connecting these two points is called the gradient-line. Thus, a line maps into a line. The perpendicular distance of the gradient-space line from the origin equals the tangent of the inclination of the original line to the image plane.

It can be shown that if the gradient-space were to be superimposed on the image-space, an image-line would be perpendicular to the corresponding gradient-space line. Mackworth's scheme for scene analysis of line-drawings of polyhedra depends on this observation [2].

TRI-HEDRAL CORNERS:

The points in gradient-space, that correspond to the three planes meeting at a tri-hedral corner, have to satisfy certain constaints. The lines connecting these points have to be perpendicular to the corresponding lines in image-space.

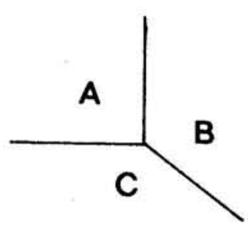
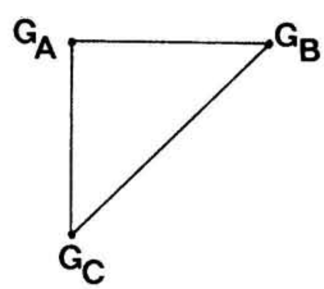



IMAGE-SPACE




GRADIENT-SPACE

This provides us with three constaints -- not enough to fix the position of three points in gradient-space. Three degrees of freedom are still undetermined, namely the position and scale of the triangle. We shall see later that measuring the three intensities provides enough information to disambiguate the orientations of the planes, and thus allows a determination of the three-dimensional structure of a polyhedral scene.

GRADIENT-SPACE IMAGE:

The amount of light reflected by a given surface element depends on its orientation and the distribution of light-sources around it, as well as on the nature of its surface. For a given type of surface and distribution of light-sources, there is a fixed value of reflectance for every orientation of the surface normal and, hence, for every point in gradient-space. Image intensity is a single-valued function of p and q. We can think of this as a gradient-space image. This is not a transform of the image seen by the viewer. It is, in fact, independent of the scene and a function of the surface properties and the light-source distribution. Note that we have assumed that both viewer and light-sources are far from the objects in the scene.

The use of the gradient-space diagram is analogous to the use of the hodogram or velocity-space diagram. The later provides insight into the motion of particles in force fields that is hard to obtain by algebraic reasoning alone. Similarly, the gradient-space will allow geometric reasoning about surface orientation and image intensities.

## MAT SURFACES AND POINT-SOURCE NEAR VIEWER:

Some examples will make this clear. Consider a perfect lambertian surface. A perfect diffuser has the property that it looks equally bright from all directions and that the amount of light reflected depends only on the cosine of the incident angle. In order to postpone the calculation of incident, emittance, and phase angles from p and q for now, we will place a single light-source near the viewer. Then the incident angle equals the emittance angle and is simply the angle between the surface normal and the view-vector. Its cosine is just the dot-product of the corresponding unit vectors. That is,

$$\cos(i) = \frac{(p,q,-1)\cdot(0,0,-1)}{|(p,q,-1)|\,|(0,0,-1)|} = 1/\sqrt{1 + p^2 + q^2}$$

The same result could have been obtained by remembering that the distance from the origin in gradient space is the tangent of the angle between the surface-normal and the view-vector:

$$\sqrt{p^2 + q^2} = \tan(e) \text{ and } \cos^2(e) = 1/[1 + \tan^2(e)] \text{ and } e = i \text{ here.}$$

If we plot reflectance as a function of p and q, we get a central maximum of one at the origin, and a circularly symmetric function that monotonically falls to zero as one goes to infinity in gradient space. This is a nice, smooth gradient-space image, typical of mat surfaces.

A given image intensity corresponds to a simple locus in gradient-space, a circle centered on the origin. A measurement of image intensity tells us that the surface gradient has to be one that falls on a certain circle in gradient-space.

UNIFORM ILLUMINATION:

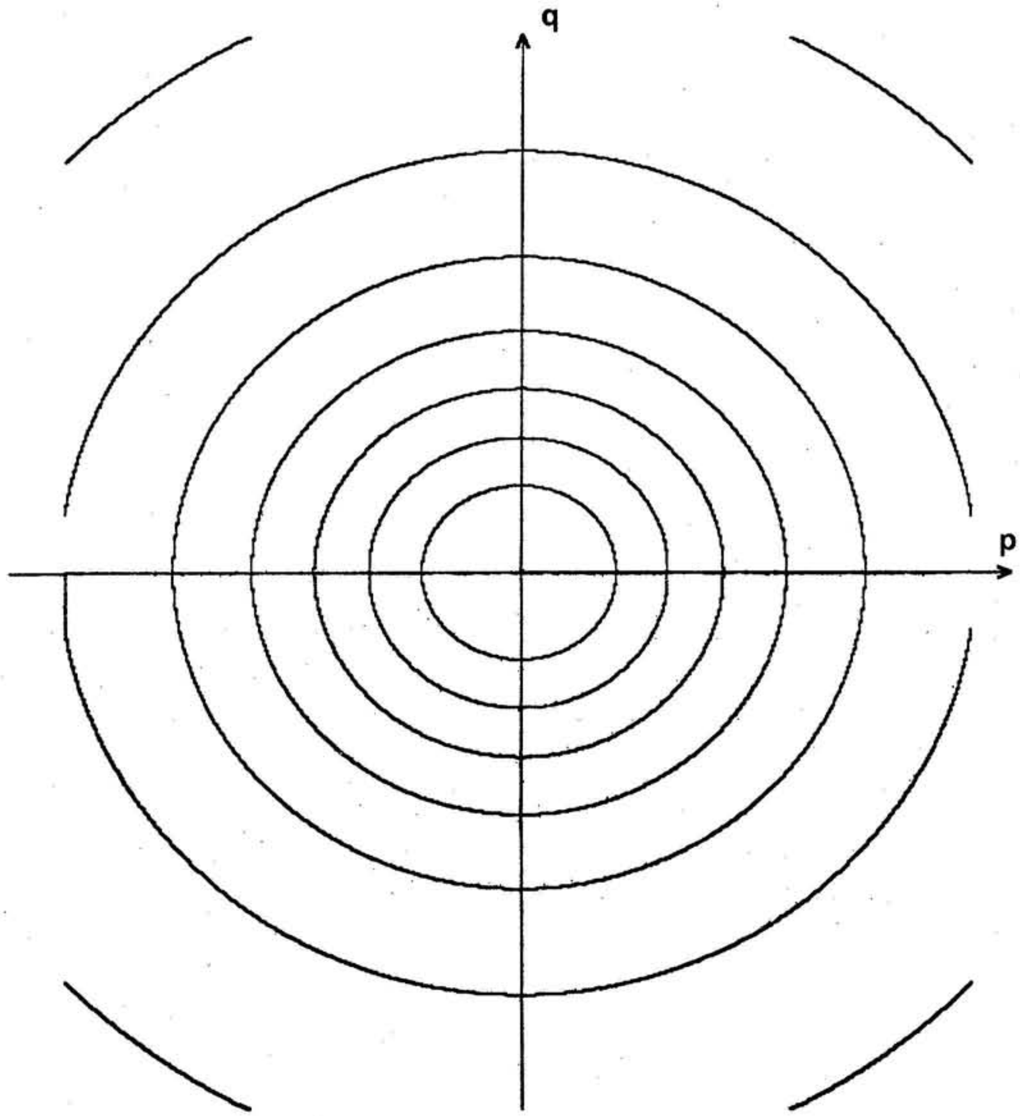
Note that the case of uniform illumination is quite similar to the situation where the light-source is near the viewer. For a start, there are no shadows in either case. Secondly, in both cases the reflectivity can be written as a function of the emittance angle alone. In fact, we can define an equivalent reflectivity function,

$$\phi'(e) = \int_{0}^{\pi/2} \int_{\pi}^{\pi} \phi(i,e,g)\ \sin(i)\ dA\,di$$

for the uniformly illuminated surface. Here A is the azimuth angle defined by

$$\cos(A) = \frac{\cos(g) - \cos(i)\cos(e)}{\sin(i)\sin(e)}$$

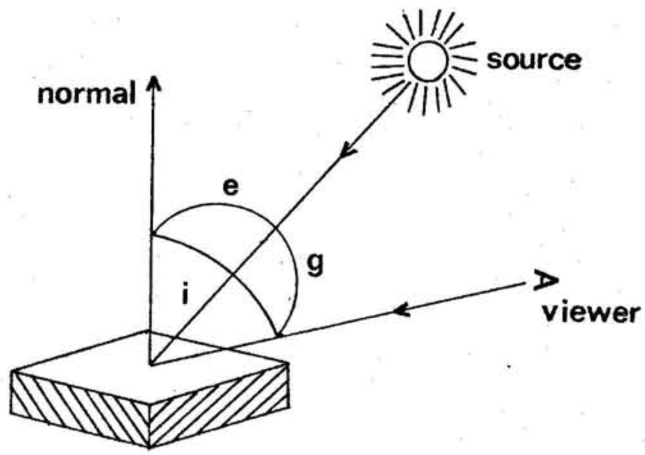
In general, the light-source is not likely to be near the viewer, so we will have to explore the more complicated geometry of incident and emitted rays for arbitrary directions of incident light at the object.




Contours of constant E = cos(e). Contour intervals are .1 units wide. This is the gradient-space image for objects with lambertian surfaces when there is a single light-source near the viewer.

INCIDENT, EMITTANCE, AND PHASE ANGLES:

For many surfaces the reflectance is a smooth function of the incident, emittance, and phase angles.




It is convenient to work with the cosines of these angles, $I = \cos(i)$, $E = \cos(e)$, and $G = \cos(g)$ -- since these can be obtained easily from dot-products of the three unit vectors. Suppose for now that we have a single distant light-source and that its direction is given by a vector $(p_s, q_s, -1)$. The view-vector is $(0,0,-1)$, so:

$$G = 1/\sqrt{1 + p_s^2 + q_s^2},\ E = 1/\sqrt{1 + p^2 + q^2},\ \text{and}$$

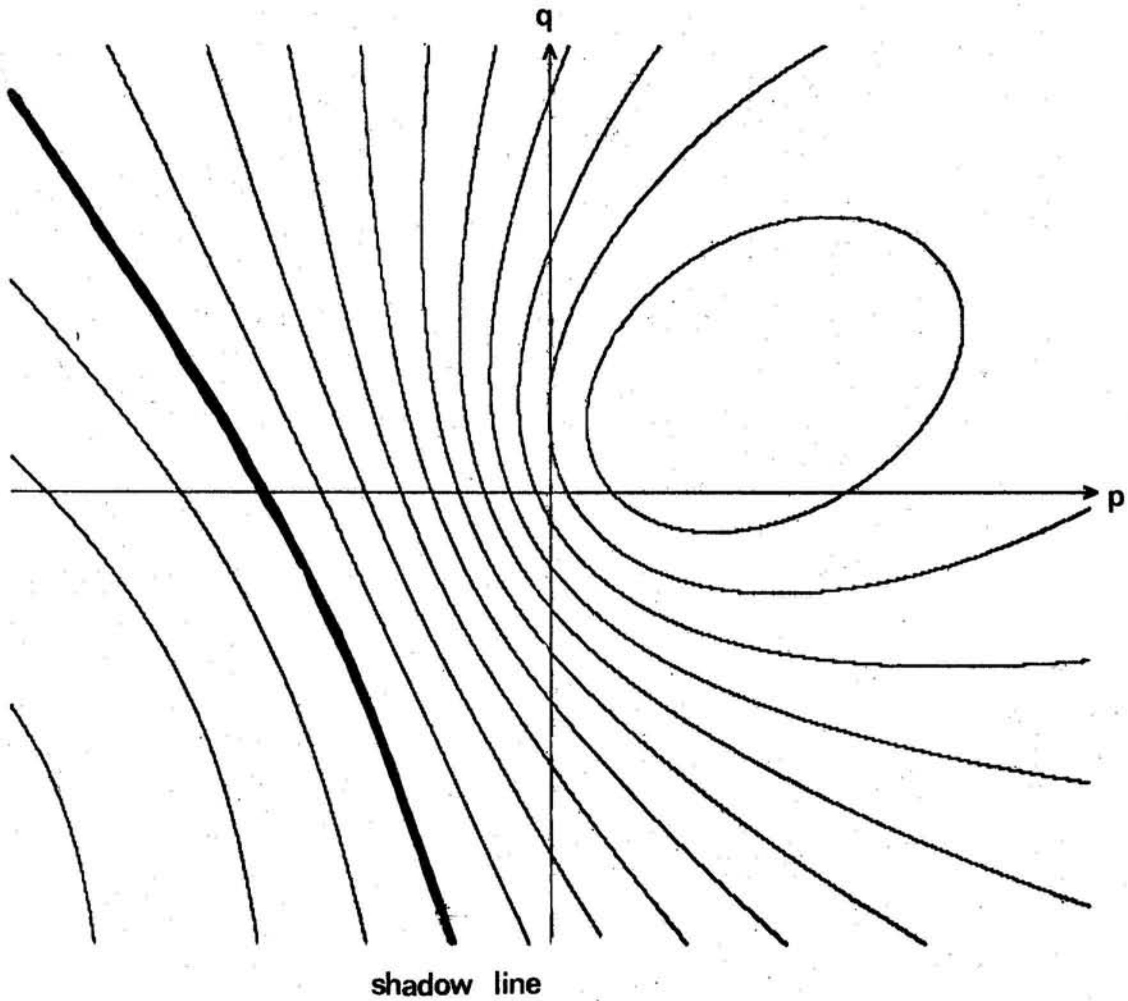
$$I = (I + p_s p + q_s q)/(\sqrt{1 + p^2 + q^2}\ \sqrt{1 + p_s^2 + q_s^2}) = (I + p_s p + q_s q)EG$$

Evidently it is simple to calculate I, E, and G for any point in gradient-space. In fact G is constant given our assumption of orthogonal projection and distant light-source. We have already seen that the contours of constant

E are circles in gradient-space centered on the origin. Setting I constant gives us a second-order polynomial in p and q and suggests that loci of constant I may be conic sections. The terminator, the line separating lighted from shadowed regions, is a straight line, obtained by setting $i = \pi/2$. Here $I = 0$; that is, $1 + p_s p + q_s q = 0$. Similarly, the locus of $I = 1$ is the single point $p = p_s$ and $q = q_s$.

A geometric way of constructing the loci of constant I is to think of the cone generated by all directions that have the same incident angle. The axis of the cone is the direction to the light-source $(p_s, q_s, -1)$. The corresponding points in gradient-space are found by intersecting this cone with a plane at unit distance from the origin. Varying values of I will produce cones with varying angles. These cones will form a nested sheaf. The intersection of this nested sheaf with the unit plane will be a nested set of conic sections.

If we measure a particular image intensity, we know that the gradient of the corresponding surface element has to fall on a particular one of the conic sections. The possible normals are then confined to a cone. In this case this is simply a circular cone. In the case of more general reflectivity functions, the locus of possible normals will constitute a more general figure called the Monge cone.



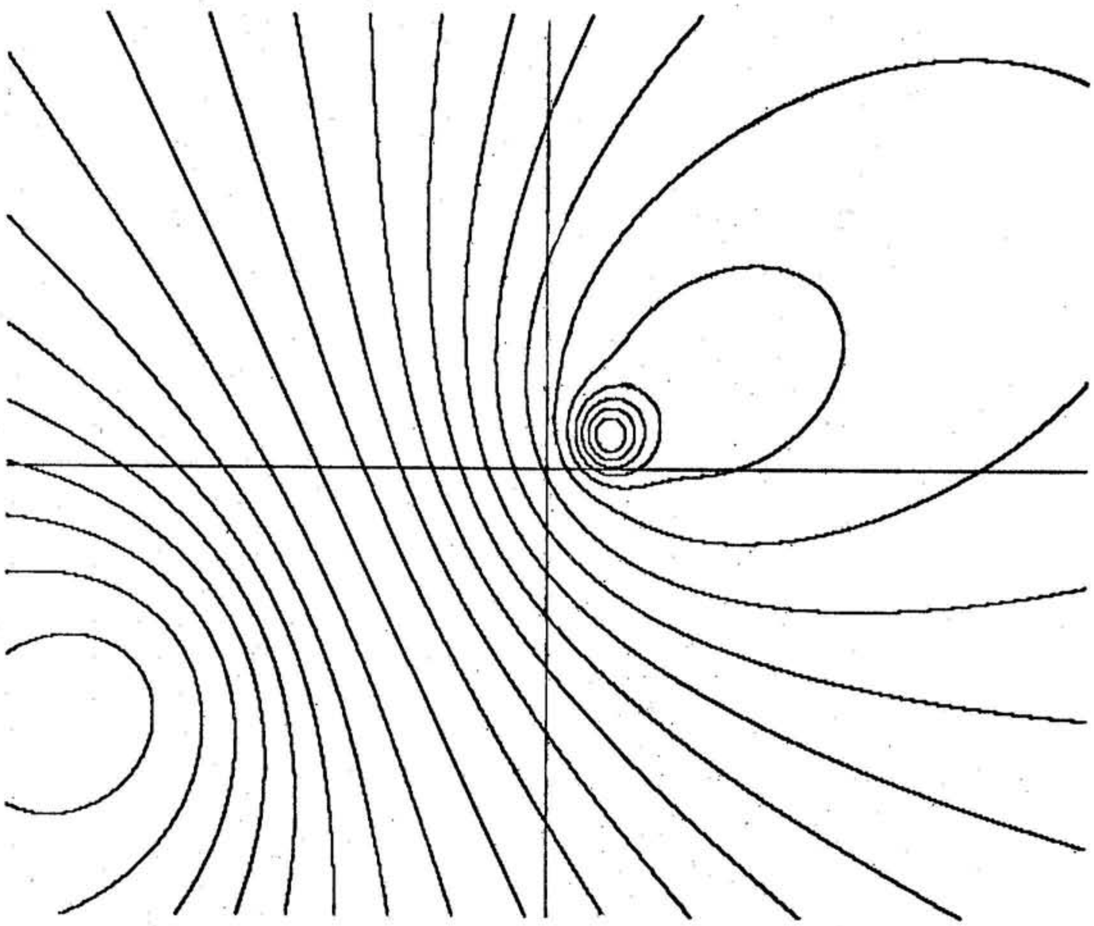

Contours of constant I = cos(i). Contour intervals are .1 units wide.
The direction to the source is $(p_s,q_s) = (0.7,0.3)$.
This is the gradient-space image for objects with lambertian surfaces when the light-source is not near the viewer.

SPECULARITY:

Many surfaces are not completely mat, having some specular reflection from the outermost layers of their surface. This is particularly true of surfaces that are smooth on a microscopic scale. For specular reflection we have $i = e$ and the incident, emitted, and normal vectors are all in the same plane. Alternatively, we can say that $i + e = g$. In any case, only one surface orientation will be just right for reflection of the light-source towards the viewer. That is, perfect specular reflection contributes an impulse to the gradient-space image at a particular point.

In practice, few surfaces have such perfect specularity. Instead they reflect some light in the direction slightly away from the geometrically correct direction [8]. It can be shown that the cosine of the angle between the direction defined by perfectly specular reflection and any other direction is $(2IE-G)$. This will clearly equal one in the correct direction and fall off towards zero as one increases the angle to a right-angle. By taking various functions of $(2IE-G)$ one can construct more or less compact specular contributions. Raising this function to some large power, for example, will do.

A good approximation for some glossy white paints can be obtained by combining the usual mat component with a specular component defined in this way. For example, $\phi(I,E,G) = \frac{1}{2}s(n + 1)(2IE - G)^n + (1 - s)I$ will work. Here s varies between 0 and 1 and determines the fraction of incident light reflected specularly before penetrating the surface, while n determines the sharpness of the specularity peak in the gradient-space image.



Contours for $\phi(I,E,G) = \frac{1}{2}s(n + 1)(2IE - G)^n + (1 - S)$ . This is the gradient-space image for a surface with both a matt and a specular component of reflectivity illuminated by a single point-source.

## FINDING p AND q FROM I, E, AND G:

In order to explore further the relation between the specification of surface orientation in gradient-space and the angles involved, we shall solve for p and q, given I, E, and G. We have already shown that it is simple to perform the opposite operation. One way of approaching this problem is to try to solve the polynomial equations in p and q derived from the equations for I, E, and G. This turns out to be messy, but it can be shown that:

$$p = p' \cos(\theta) - q' \sin(\theta)$$

$$q = p' \sin(\theta) + q' \cos(\theta)$$

Where

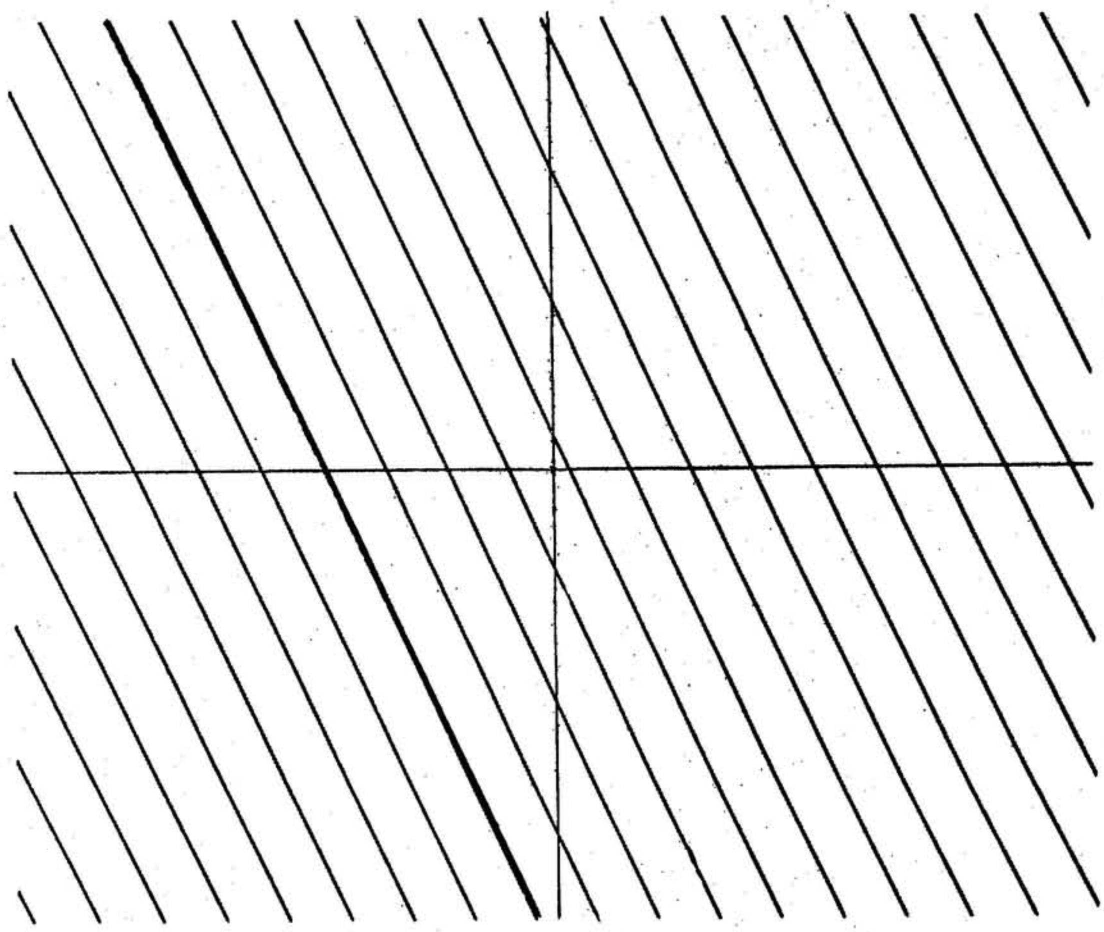
$$p' = \frac{(I/E - G)}{\sqrt{1 - G^2}} \quad \text{and} \quad q' = \frac{\pm(\Delta/E)}{\sqrt{1 - G^2}}$$

$$\Delta^2 = 1 + 2IEG = (I^2 + E^2 + G^2)$$

$$\cos(\theta) = \frac{p_s}{\sqrt{p_s^2 + q_s^2}} \quad \text{and} \quad \sin(\theta) = \frac{q_s}{\sqrt{p_s^2 + q_s^2}}$$

It is immediately apparent that for most values of I, E, and G, there are two solution points in gradient space. Notice that $\theta$ here is the direction of the light-source in gradient-space; the line connecting $(p_s, q_s)$ to the origin makes an angle $\theta$ with the p-axis. So p' and q' are coordinates in a new gradient-space obtained after simplifying matters by rotating the axes until $q_s = 0$ -- the light source is in the direction of the x'-axis.

The next thing worth noticing about this set of equations is that if I/E is constant, then p' is constant (remembering that G is constant anyway). So the loci of constant I/E are straight lines. These lines are all parallel to the terminator, for which I = 0. This turns out to be important since some surfaces have constant reflectance for constant I/E.



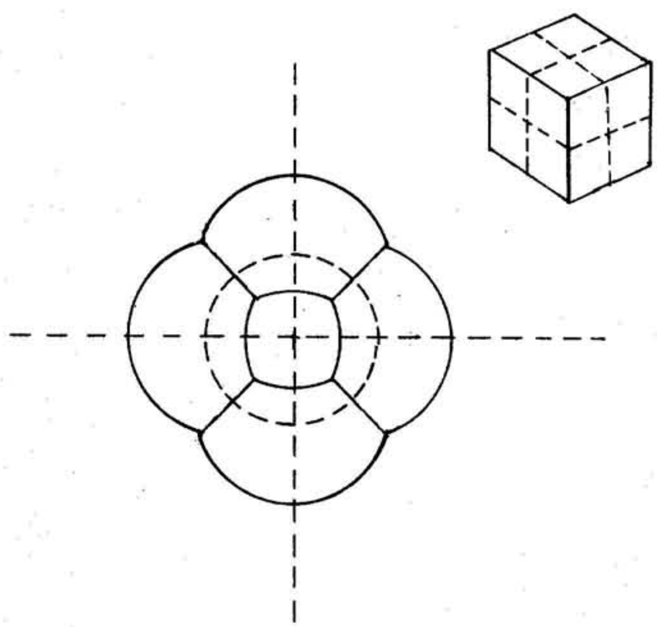
Contours of $\phi(I,E,G) = I/E$. Contour intervals are .2 units wide. The reflectivity function for the material in the maria of the moon is constant for I/E.

METALLIC SURFACES:

Consider next a metallic surface, a surface with a purely specular reflectance. Each point in gradient-space corresponds to a particular direction of the surface normal and defines a direction from which incident light has to approach the object in order to be reflected towards the viewer. In fact, in gradient-space we can produce a complete map of the sphere of possible directions as seen from the object. At the origin, for example, we have the direction towards the viewer. If we record an intensity in the gradient space corresponding to the intensity arriving at the object from the corresponding direction we obtain a picture of the world surrounding the object. In map projection terms we have a plane projection of a sphere with one pole of the sphere as the center of projection. Another way of looking at it is that the image we construct in this fashion is like one we would obtain by looking into a convex mirror -- a metallic paraboloid to be precise.

What can we do with this strange image of the world surrounding the object? If we measure a certain intensity at a given point on the object, we can now say something about the orientation of the surface at that point. We cannot uniquely determine that orientation, but we do know that it is restricted to a sub-set of all possible orientations. We have one constraint on it -- it has to be one of the points in gradient-space where we find this same value of intensity. If the world surrounding the object is at all complex, this sub-set will tend to be very disconnected and complex, and not much help in recovering the shape directly. There are exceptions -- light-sources, for example, tend to be compact and very bright, corresponding to definite easy-to-locate points in gradient-space. For points with such high reflected light intensity in the image we can often locally determine the surface normal uniquely.

We have now developed methods for constructing gradient-space images for various surfaces and distributions of light-sources. The latter is done simply by superimposing the results in gradient-space for each light-source in turn. We will now turn to a minor flaw in this approach and attempt a partial analysis of mutual illumination.



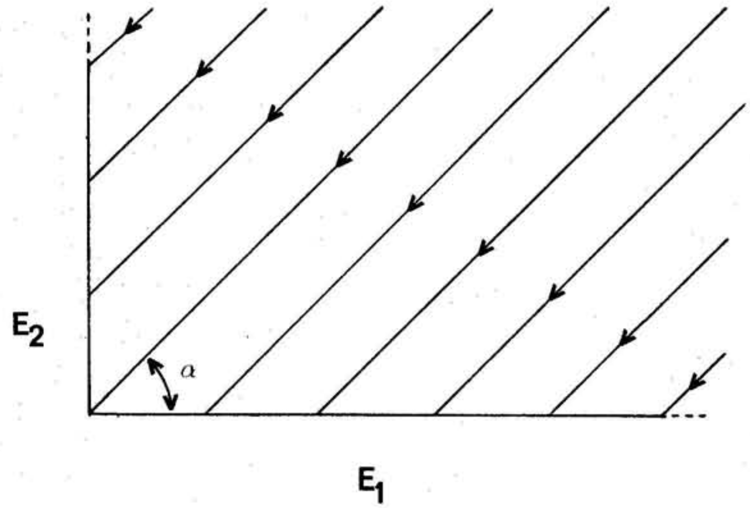
Gradient-space image for a metallic object in the center of a large wire cube. Equivalently one can think of it as the reflection of the wire cube in a paraboloid with a specularly-reflecting surface.

## MUTUAL ILLUMINATION:

The gradient-space image is based on the assumption that the viewer and all light sources are distant from the object. Only under these assumptions can we associate a unique value of image intensity with every surface orientation. If the scene consists of a single convex object these assumptions may be satisfied, but when there are several highly reflective objects placed near one another, mutual illumination may become important. That is, the distribution of incident light no longer depends only on direction only, but is a function of position as well. The general case is very difficult to deal with and we shall study only some idealized situations applicable to scenes made up of polyhedra. There are two primary effects of mutual illumination: a reduction in contrast between faces, and the appearance of shading or gradation of light on images of plane surfaces. In the absence of this effect, we would expect plane surfaces to have polygonal images of uniform intensity since all points on them have the same orientation.

## TWO SEMI-INFINITE PLANES:

First let us consider a highly idealized situation where we have two semi-infinite planes joined at right angles, and a distant light-source. Let the incident rays make an angle $\alpha$ with respect to one of the planes. Further assume that the surfaces reflect a fraction r of the light falling on them, and that the illumination provided by the source is E (light flux/unit area).




Picking any point on one of the half-planes, we find that one-half of its hemisphere of directions is occupied by the other plane, so one-half of the light radiated from this point will hit the other plane, while one-half will be lost. Since both planes are semi-infinite, the geometry of this does not depend on how far from the corner we are. Now, the light incident at any point is made up of two components, that received directly from the source and that reflected from the other plane. It is not hard to see that the intensity on one plane will not vary with distance from the corner -- a point receives reflected light from one-half of its hemisphere of directions no matter how far from the corner it is. Put another way, there is no natural scale factor for a fluctuation in intensity. Let the illumination of the planes be $E_1$ and $E_2$ (light flux/unit area).

$$E_1 = \tfrac{1}{2}E_2 + E\cos(\alpha)$$

$$E_2 = \tfrac{1}{2}E_1 + E\sin(\alpha)$$

Solving for $E_1$ and $E_2$, one gets:

$$E_1 = E[\cos(\alpha) + \tfrac{1}{2}r \sin(\alpha)]/[1 - (\tfrac{1}{2}r)^2]$$

$$E_2 = E[\sin(\alpha) + \tfrac{1}{2}r \cos(\alpha)]/[1 - (\tfrac{1}{2}r)^2]$$

Had we ignored the effects of mutual illimination we would have found $E_1 = E \cos(\alpha)$ and $E_2 = E \sin(\alpha)$. Clearly the effect increases with increases in reflectance r; it is not significant for dark surfaces. When the planes are equally illuminated, for $\alpha = \pi/4$, we have:

$$E_1 = E_2 = (E/\sqrt{2})/(1 - \tfrac{1}{2}r)$$

When r = 1, this is twice the illumination and hence twice the brightness that we would have obtained in the absence of mutual illumination.

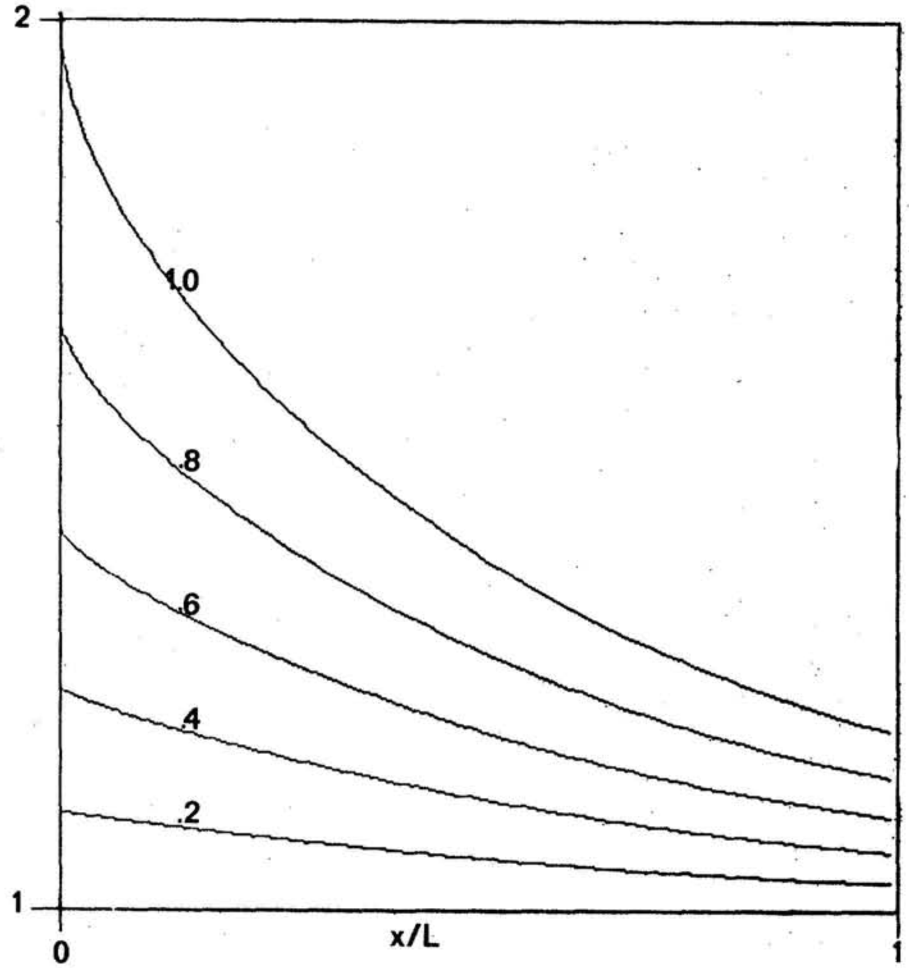
If the angle between the two planes is varied, one finds that the effect gets larger and larger as the angle gets more and more acute. One can get arbitrary "amplification" by choosing the angle small enough. Conversely, for angles larger than $\pi/2$, the effect is less pronounced.

In the above derivation we have not made very specific assumptions about the angular distribution of reflected light, just that it does not depend on where the incident ray comes from and that it is symmetrical about the normal. So a lambertian surface would be included, while a highly specular one would not. Indeed, the effect is less pronounced for surfaces with a high specular component of reflection, since most of the light is bounced back at the source upon second reflection.

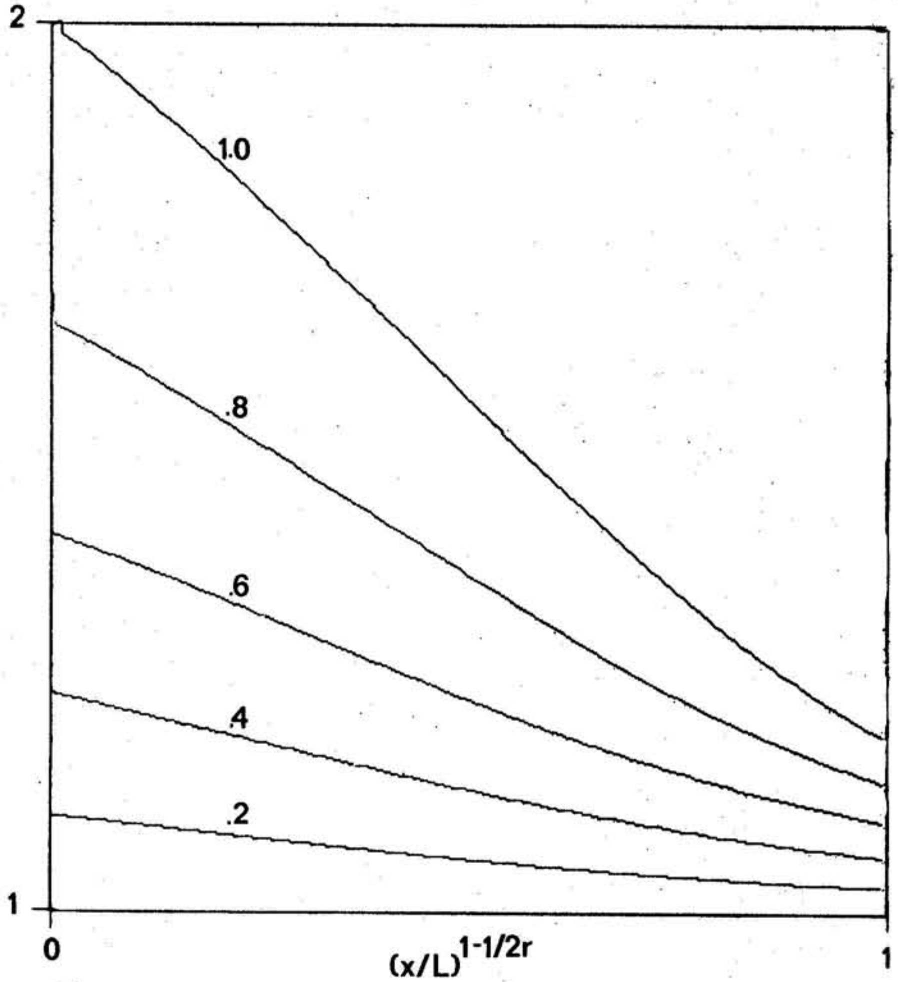
Another important thing to note is that if the planes are not infinite, the above calculations apply approximately at least close to the corner. For finite planes we expect a variation of intensity as a function of distance from the corner, but asymptotically, as one approaches the corner, the results derived here will apply.

## TWO TRUNCATED PLANES:

If the planes are of finite extent, the geometry becomes quite complex, but, if one allows them to be infinite along their line of intersection and truncates them only in the direction perpendicular to this, one can develop an integral equation. Suppose they both extend a distance L from the corner, and are joined at right-angles and that $\alpha = \pi/4$. This produces a particularly simple form of this integral equation -- which nevertheless I have been unable to solve analytically. Numerical methods show that the resultant illumination falls off monotonically from the corner, that the value at the corner is indeed what we predicted in the previous section, and that near the corner, the fall-off is governed by a term in $-(x/L)^{(1 - \frac{1}{2}r)}$. For $r = 1$, for example, this contains the square-root of $(x/L)$ and there is thus a cusp in the function at the corner. (Here x is the distance along the plane from the edge where the planes meet).




Surface luminance plotted versus fractional distance from a right-angle corner. The curves are for reflectances of .2, .4, .6, .8, and 1.0.

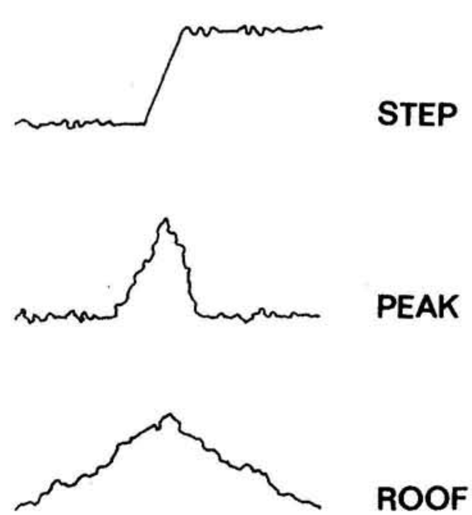



Surface luminance plotted versus $(x/L)^{(1-\frac{1}{2}r)}$ to illustrate asymptotic behavior near the corner. The curves correspond to reflectances of .2, .4, .6, .8, and 1.0.

# THE MAIN RESULTS

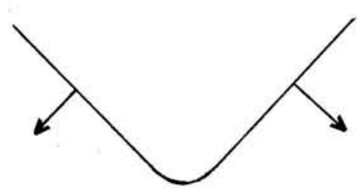
## THE SEMANTICS OF EDGE-PROFILES:

If polyhedral objects were perfect, there was no mutual illumination, image sensors were perfect and light sources distant from the scene, images of polyhedral objects would be divided into polygonal areas, with intensity uniform inside each polygon. It is well known that there is variation of image intensity within these polygonal areas in real images and that an intensity profile taken across an edge separating two such polygonal regions does not simply have a step-shaped transition in intensity. Herskovitz and Binford determined experimentally that the most common edge transitions are step-, peak-, and roof-shaped [7]. This has so far been considered no more than a nuisance, since it complicates the process of finding edges. Here we will discuss the interpretation of these profiles in terms of the three-dimensional aspects of the scene.



IMPERFECTIONS OF POLYHEDRAL EDGES:

A perfect polyhedron has a discontinuity in surface normal at an edge. In practice edges are rounded off somewhat. A cross-section through the object's edge shows that the surface normal varies smoothly from one value to the other and takes on values that are linear combinations of the surface normals of the two adjoining planes.



What does this mean in terms of reflected light intensity? Instead of a sudden jump of intensity from a value corresponding to the one surface normal to the other, the intensity varies smoothly. The important point is that it may take on values outside the range of values defined by the two planes. The best way to see this is to consider the situation in gradient-space. The two planes define two points in gradient-space and tangent planes on the corner correspond to points on the line connecting these two points. If the image intensity is higher for a point somewhere on this line, we will see a peak in the intensity profile across the edge.

So, if we find an edge-profile with a peak-shape or a step with a peak superimposed, it is most likely that the line should be labelled convex. The converse is not true, an edge may be convex and not give rise to a peak, if the line connecting the two points in gradient space has intensity varying monotonically along its length. The identification is also not completely certain since under peculiar lighting conditions and with objects that have acute angles between adjacent faces, a peak may appear at an obscuring edge.

Notice that the peak is quite compact, since it only extends as far as the rounded-off edge does.

At a corner, where the planes meet, we find that surface imperfections provide surface normals that are linear combinations of the three normals corresponding to the three planes. In gradient space this corresponds to points in the triangle connecting the three points corresponding to the planes. If this triangle contains a maximum in image intensity we expect to see a high-light right on the corner.

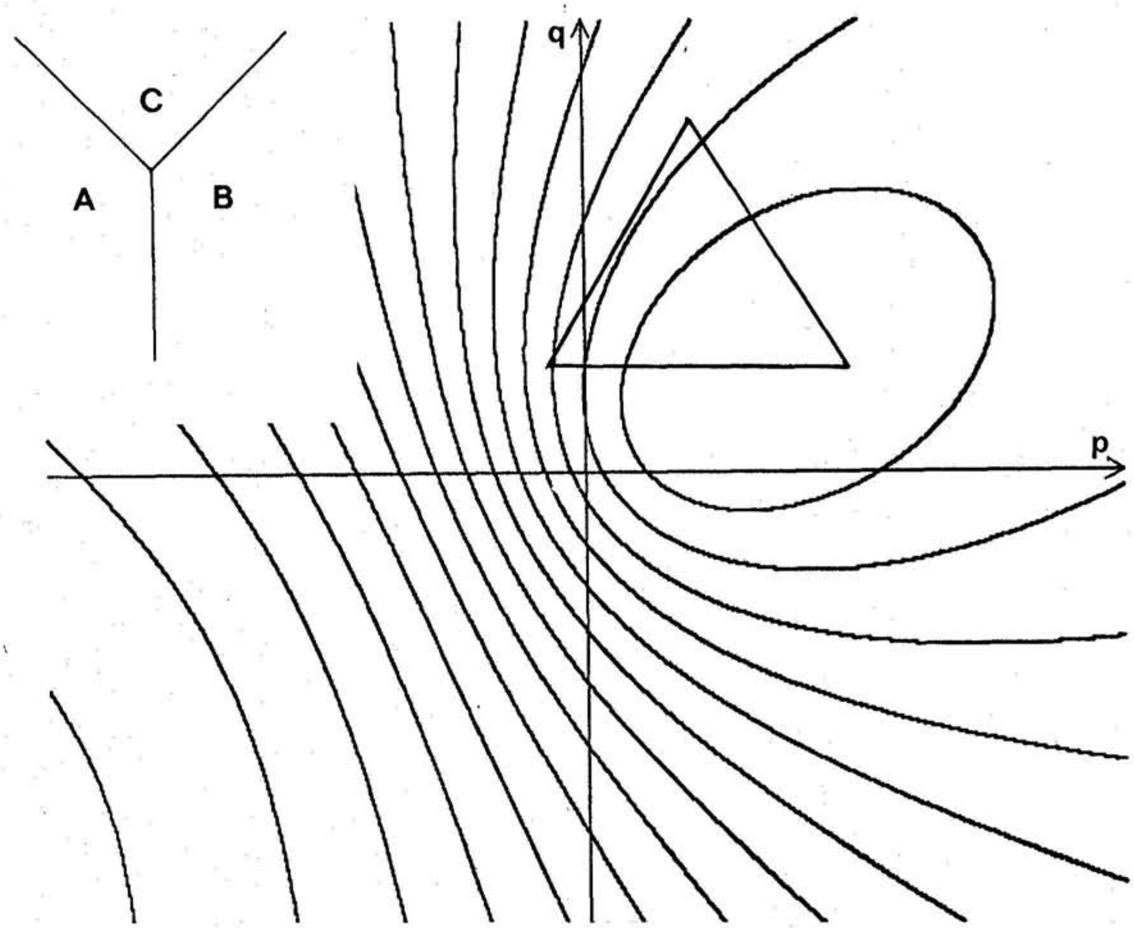



Image of tri-hedral corner and corresponding gradient-space diagram. The image intensity profile across the edge between face A and face B will have a peak or highlight. The others will not.

MUTUAL ILLUMINATION:

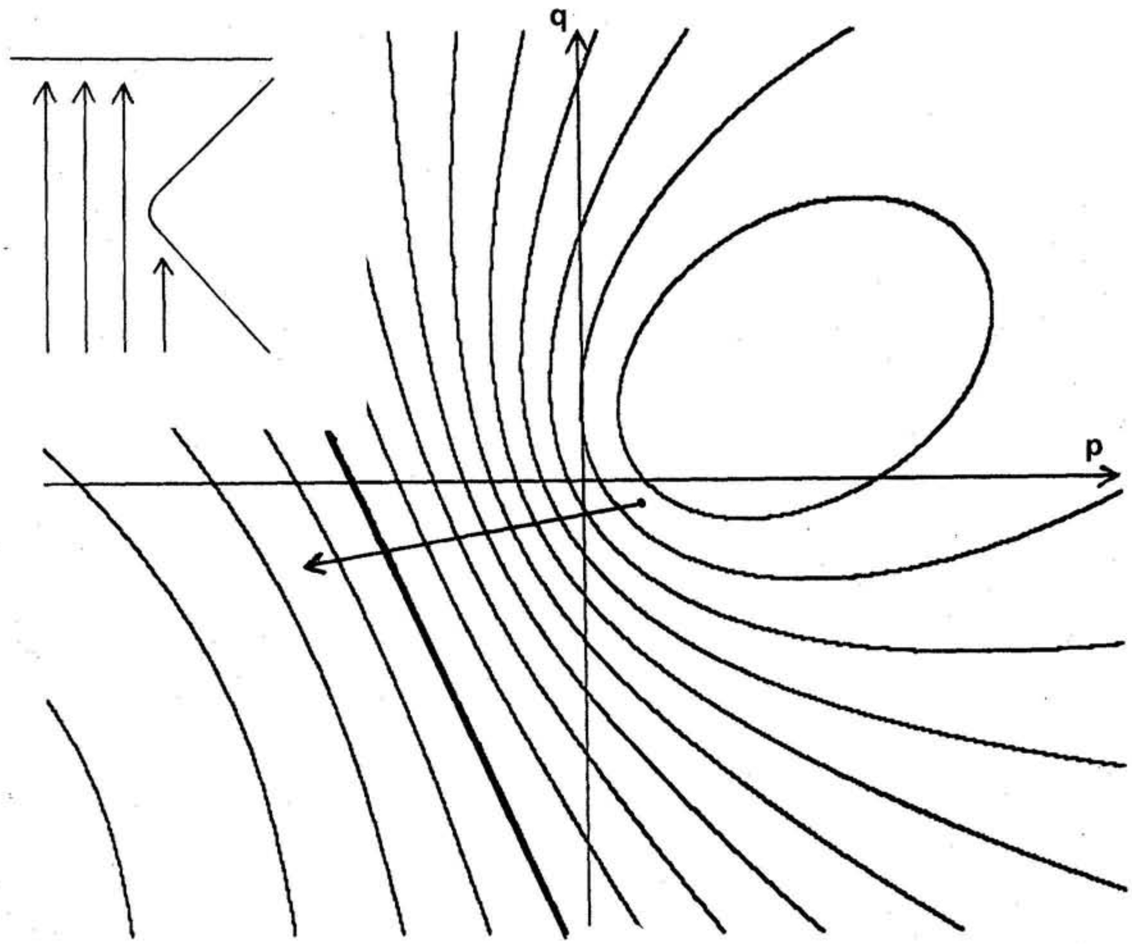
We have already seen that mutual illumination gives rise to intensity variations on planar surfaces. The intensity falls off as one moves away from the corner. Near the corner, this fall-off is approximately linear. Notice that this affects the intensity profile over a large distance from the edge, quite unlike the sharp peak found due to edge imperfections. Clearly, if we find a roof-shaped profile or step with a roof-shape superimposed we should consider labelling the edge concave.

The identification is not perfectly certain, though, since some imaging device defects can produce a similar effect. Image dissectors, for example, suffer from a great deal of scattering and this has the effect that areas further from a dark background are brighter. So one may see a smoothed version of a roof-shape in the middle of a bright scene against a dark background. Experimentation with high-quality image input devices such as the PIN-diode mirror-deflection system has confirmed that this is an artifact introduced by the image dissector.

Further, when the light-source is close to the scene, significant gradients can appear on planar surfaces as pointed out by Herskovitz & Binford [7]. Lastly, the roof-shaped profiles on the two surface may be due to mutual illumination with other surfaces, not each other. Nevertheless, a roof-shaped profile does usually suggest a concave edge.

OBSCURATION:

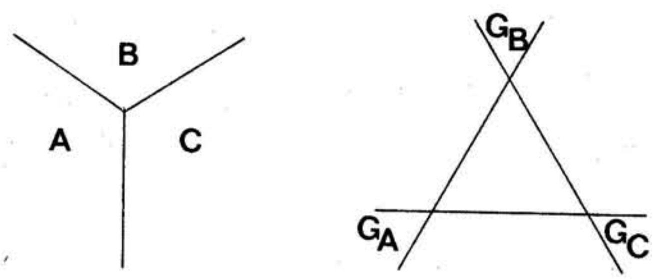
Step-shaped intensity profiles most often occur where objects obscure one another, although they can be found with convex and sometimes concave edges as well. If the obscuring surface adjoins a self-shadowed surface, however, edge imperfections will produce a negative peak on the profile, since the line connecting the points corresponding to the two surfaces in gradient-space then passes through the terminator. So a negative peak or a step with a superimposed negative peak strongly suggests obscuration. It is unfortunately impossible to tell which side is the obscuring plane.




Generation of a negative peak at an obscuring edge facing away from the light-source.

DETERMINING THE THREE-DIMENSIONAL STRUCTURE OF POLYHEDRAL SCENES:

The approach invented by Mackworth for understanding line-drawings of polyhedra allows one to take into account some of the quantitative aspects of the three-dimensional geometry of scenes [2]. It does not, however, allow one to determine fully the orientation of all the planes. The scale and position of the gradient-space diagram is undetermined by his technique. To illustrate, consider a single trihedral corner. Here we know that the three points in gradient-space that represent the three planes meeting at the corner have to satisfy certain constraints. Specifically, they must lie on three lines perpendicular to the image-lines.




It takes six parameters to specify the position of three points on a plane, so we still have three degrees of freedom after introducing these constraints. Measuring the three image intensities of the planes supplies another three. The constraints are due to the fact that the points in gradient-space have to lie on the right contours of image intensity. The triangle can be stretched and moved until the points correspond to the correct image intensities as measured for the three planes. Since this process corresponds to solving three non-linear equations for three unknowns, we can expect a finite number of solutions. Often there are but one or two; some can be eliminated from prior knowledge of what is to be expected in the scene.

When more than three planes meet at a corner, the situation is even more constrained -- the equations are over-determined. Conversely, one cannot do much with just two planes meeting at an edge, since there are too few equations, and an infinite number of solutions exist, as one might expect.

The possible ambiguity at a tri-hedral corner is not very serious when one considers that in a typical scene there will be many "connect" edges, either convex or concave as determined by Mackworth's program. Usually the overall constraints will allow only one interpretation that is consistent. A practical difficulty is that it is unclear what search strategy will lead one efficiently to this interpretation.

Measurements of image intensity are not very precise and surfaces have properties that vary from point to point and with handling. We cannot expect this method to be extremely accurate in pinning down surface orientation. The fact that for a typical scene the equations will be over-determined allows a least-squares approach which may help to improve matters a little.

The idea of stretching and shifting can be generalized to smooth surfaces. We know that the image of a paraboloid is the gradient-space image. If we can stretch and shift a real image of some object to fit this pattern of intensity distribution we can determine its surface shape by applying the inverse stretching and shifting to the paraboloid.

LUNAR TOPOGRAPHY:

When viewed from a great distance, the material in the maria of the moon has a particularly interesting reflectivity function. First, note that the lunar phase is the angle at the moon between the light-source (sun) and the viewer (earth). This is obviously the angle we call g, and explains why we use the term phase angle for g. For constant phase angle, detailed measurements using surface elements, whose projected area as seen from the source is a constant multiple of the projected area as seen by the viewer, have shown that all such surface elements have the same reflectance. But the area appears foreshortened by cos(i) and cos(e) as seen by the source and the viewer respectively. Hence the reflectivity function is constant for constant cos(i)/cos(e) = I/E (for fixed G).

Each surface element scatters light uniformly into its hemisphere of directions, quite unlike the lambertian surface, which favors directions normal to its surface. This is not an isolated incident. The surfaces of other rocky, dusty objects when viewed from great distances appear to have similar properties. The surface of the planet Mercury, for example, and perhaps Mars, as well as some asteroids and atmosphere-free satellites fit this pattern. Surfaces with reflectance a function of I/E thus form third species we should add to mat surfaces where the reflectance is a function of I and glossy surfaces where the reflectance is a function of (2IE-G).

LUNAR REFLECTIVITY FUNCTION:

Returning to the lunar surface, we find an early formula due to Lommel-Seelinger [6]:

$$\phi(I,E,G) = \frac{\Gamma_o(I/E)}{(I/E) + \lambda(G)}$$

Here $\Gamma_o$ is a constant and the function $\lambda(G)$ is defined by an empirically determined table. A somewhat more satisfactory fit to the data is provided by a formula of Fesenkov's [6]:

$$\phi(I,E,G) = \frac{\Gamma_o(I/E)[1 + \cos^2(\alpha/2)]}{(I/E) + \lambda_o[1 + \tan^2(\lambda/2)]}$$

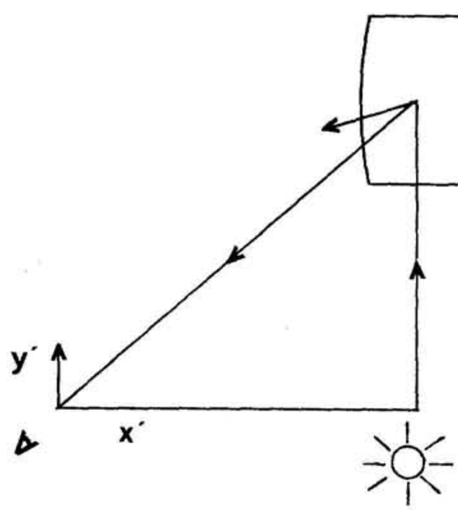
Where $\Gamma_o$ and $\lambda_o$ are constants and $\tan(\alpha) = -(I/E-G)/\sqrt{1-G^2}$. By the way, $\tan(\alpha) = -p'$. This formula is also supported by a theoretical model of the surface due to Hapke. Note that given I, E, and G, it is straightforward to calculated the expected reflectance. We need to go in the reverse direction and solve for I/E given G and the reflectance as measured by the image intensity. While it may be hard to invert the above equation analytically, it should be clear that by some iterative, interpolation, or hill-climbing scheme, one can solve for I/E. We shall ignore for now the ambiguities that arise if there is more than one solution.

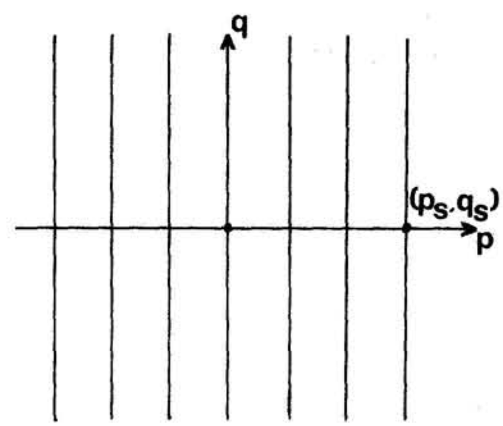
LUNAR GRADIENT-SPACE IMAGE:

Next, we ask what the gradient-space image looks like for the lunar surface illuminated by a single point-source. The contours of constant intensity in gradient-space will be lines of constant I/E. But the contours of constant I/E are straight lines! So the gradient-space image can be generated from a single curve by shifting it along a straight-line -- the shadow-line, for example. The contour lines are perpendicular to the direction defined by the position of the source (that is, the line from the origin to $p_s, q_s$).

Now what information does a single measurement of image intensity provide? It tells us that the gradient has to be on a particular straight line. Again, we shall ignore for the moment the possible existence of more than one contour for a given intensity. What we would like to know of course is the orientation of the surface element. We cannot determine completely that locally, but we can tell what its component will be in one direction, the direction perpendicular to the contour lines. We can tell nothing about it in the direction at right-angles to this favored direction. In fact, knowing I/E and G determines p', as previously defined and tells up nothing about q'.

This favored direction lies in the plane defined by the source, the viewer, and the surface element under consideration. If one wishes, one can simplify matters by rotating the viewer's coordinate system system until the x axis lies in this plane as well. Then $q_s = 0$, and the contours of constant intensity in gradient-space are all vertical lines. Evidently, an image intensity measurement determines the slope of the surface in the x' direction, without telling us anything about the slope in the y' direction.




We are now ready to integrate out the surface by advancing in the direction in which we can locally determine the surface slope.

FINDING A SURFACE PROFILE BY INTEGRATION:

We have:

$$p' = \frac{dz}{ds} = \frac{I/E - G}{\sqrt{1 - G^2}}$$

The distance s from some starting point is measured in the object coordinate system and is related to the distance along the projection of this curve in the image by $s' = s(f/z_o)$.

$$\frac{dz}{ds'} = \frac{f}{z_o} \frac{I/E - G}{\sqrt{1 - G^2}}$$

Integrating, we get:

$$z(s') = z_o + \frac{f}{z_o} \int_o^{s'} \frac{I/E - G}{\sqrt{1 - G^2}} \, ds'$$

Where I/E is found from G and the image intensity $b(x',y')$ by $\Psi_G$:

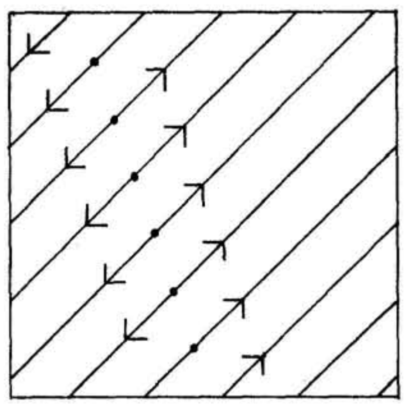
$$I/E = \Psi_G[b(x',y')]$$

Starting anywhere in the image, we can integrate along a particular line and find the relative elevation of the corresponding points on the object.

The curves traced out on the object in this fashion are called characteristics, their projection in the image plane are called base characteristics. It is clear that the base characteristics here are parallel straight lines in the image, independent of the object's shape.

## FINDING THE WHOLE SURFACE:

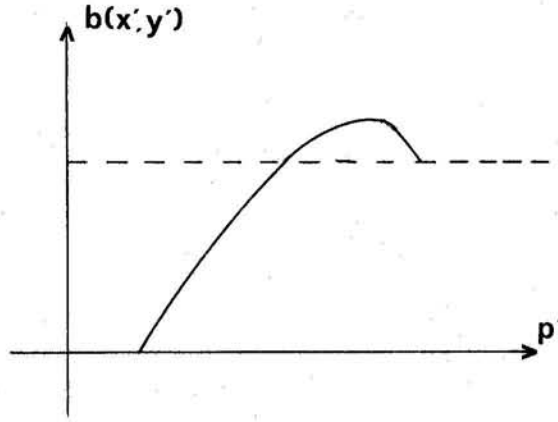
We can explore the whole image by choosing sufficient starting points along a line at an angle to the favored direction. In this way we obtain the surface shape over the whole area recorded in the image.



There is nothing to relate the integrals obtained along adjacent characteristics in the image, since we cannot determine the gradient in this direction. We have to know an initial curve, or use assumptions of reasonable smoothness. Alternatively, we can perform a second surface calculation from an image taken with a different source-surface-observer geometry. In this case, we will obtain solutions along lines crossing the surface at a different angle and can so tie the two solutions together. This is not quite as useful as one might think at first, since it does not apply to pictures taken from earth. The plane of the sun, moon, and earth varies little from the ecliptic plane. The lines of integration in the image will vary little in inclination. This idea does work for pictures taken close to the moon. to the moon.

AMBIGUITY IN LOCAL GRADIENT:

What if more than one contour in gradient-space corresponds to a given intensity? Then we cannot tell locally which gradient to apply. If we are integrating along some curve, however, this is not a problem, since we may assume that there is little change in gradient over small distances and pick the one close to the gradient last used. This assumption of smoothness leaves us with one remaining problem: what happens if we approach a maximum of intensity in gradient-space and then enter areas of lower intensity. Which side of the local maximum do we slide down? This is an ambiguity which cannot be resolved locally, and the solution has to be terminated at this point. Under certain lighting conditions the image will be divided into regions inside each of which we can find a solution. The regions will be separated by ambiguity edges, which cannot be crossed without making an arbitrary choice.



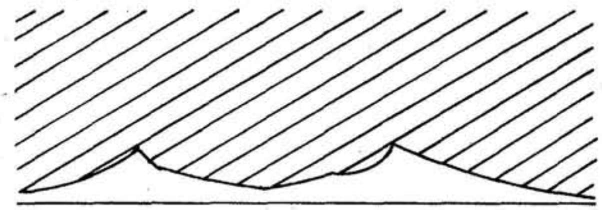
LOW SUN-ANGLES:

This problem can be entirely avoided if one deals only with pictures taken at low sun-angles, since the gradient is then a single valued function of image intensity. This is a good idea in any case, since the accuracy of the reconstruction will depend on how accurately one can determine the gradient, which in turn depends on the spacing of the contour lines in gradient-space. If they are close together, this accuracy will be high; near a maximum, on the other hand, it will be low. It is easy to convince oneself that pictures taken at low sun-angle have "better contrast," show the "relief in more detail", and are "easier to interpret".

There is another reason for interest in images obtained under conditions of low sun-angle. Near the shadow-line in gradient-space, the contours of constant-intensity are nearly straight lines even if we are not dealing with the special reflectivity function for the lunar material! An early solution to the problem of determining the shape of lunar hills made use of this fact by integrating along lines perpendicular to the terminator [5].

DEALING WITH SHADOWS:

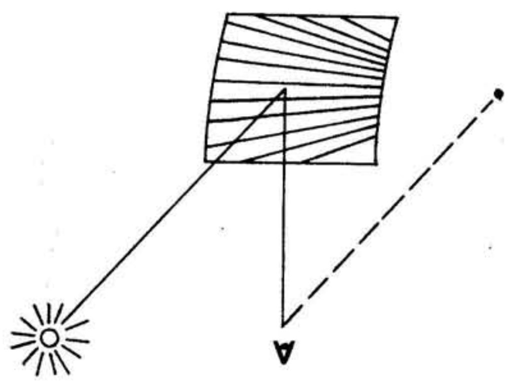
Working at low sun-angles introduces another problem of course, since shadows are likely to appear. Fortunately, they are easy to deal with since we can simply trace the line in the image until we again see a lighted area. Since we know the direction of the rays from the source we can easily determine the position of the first lighted point. The integration is then continued from there. In fact, no special attention has to be paid to this problem, since a surface element oriented for grazing incidence of light will already have the correct slope. Thus simply looking up the slope for zero intensity and integrating with this value will do.



Some portion of the surface of course will not be explored because of shadows. Most of this area will be covered if one takes one picture just after "sun-rise" and one just before "sun-set".

GENERALIZATION TO PERSPECTIVE PROJECTION:

All along we have assumed orthographic projection -- looking at the surface from a great distance with a telephoto lens. In practice, this is an unreasonable assumption for pictures taken by artificial satellites near the surface. The first thing that changes in the more general case of perspective projection is that the sun-surface-viewer plane is no longer the same for all portions of the surface imaged. Since it is this plane which determines the lines along which we integrate, we can expect that the lines of integration will no longer be parallel. Instead they all converge on the anti-solar point -- that is, the point in the image which corresponds to a direction directly opposite to the direction towards the source.



The next change is that z is no longer constant in the projection equation. So $s' = f(s/z)$. Hence,

$$p' = \frac{dz}{ds} = \frac{f}{z}\frac{dz}{ds'} = \frac{1/E - G}{\sqrt{1 - G^2}}$$

We can no longer simply integrate. But it is easy to solve the above differential equation for z by separating terms:

$$\log(z) = \frac{1}{f}\int \frac{1/E - G}{1 - G^2}\, ds' \text{ and so } z(s') = z_o\, e^{\frac{1}{f}\int_0^{s'} \frac{1/E - G}{\sqrt{1 - G^2}}\, ds'}$$

Finally, note that the phase angle g is no longer constant. This has to be taken into account when calculating I/E from the measured image intensity. On the whole, the process is still very simple. The paths of integration are pre-determined straight lines in the image -- radiating from the anti-solar point. At each point we measure the image intensity, determine what value of I/E will give rise to this image intensity. Then we calculate the corresponding slope along the straight line and take a small step. Repeating for all lines crossing the image we obtain the surface elevation at all points in the image.

The same result could have been obtained by a very painful algebraic method [6].

A NOTE ON ACCURACY:

Since image intensities can only be determined with rather limited precision, one must expect the calculation of surface coordinates to suffer from errors that may accumulate along characteristics. A "sharpening" method that relates adjacent characteristics can reduce these errors somewhat [4]. It further appears that an objects shape is better described by the orientations of its surface normals than by distances from the viewer to points on its surface. In part this may be because distances to the surface undergo a more complicated transformation when the object is rotated than do surface normal directions. Note that the calculation of surface normals is not subject to the cumulative errors mentioned.

Finally, it should be pointed out that the precise determination of the surface shape is not the main impetus for the devlopment presented here. The understanding of how image intensities are determined by the object, the lighting and the image forming system is of more importance and may lead to interesting heuristic methods.

GENERAL REFLECTIVITY FUNCTIONS:

The simple method developed for lunar topography does not apply if the contours of constant intensity in gradient-space are not parallel straight lines. We shall still be able to trace along the surface, but the direction we take at each point will now depend on the image and will change along the profile. The base characteristics will no longer be pre-determined straight lines in the image. At each point on a characteristic curve we shall find that the solution can be continued only in a particular direction. It will also appear that we will need more information to start a solution and shall have to carry along more information as we proceed. Reasoning from the gradient-space diagram can be augmented here by some algebraic manipulation.

Let $a(p,q)$ be the intensity corresponding to a surface element with a gradient $(p,q)$. Let $b(x,y)$ be the intensity recorded in the image at the point $(x,y)$. Then, for a particular surface element, we must have:

$$a(p,q) = b(x,y)$$

Now suppose we want to proceed in a manner analogous to the method developed earlier by taking a small step $(dx,dy)$ in the image. It is clear that we can calculate the corresponding change in z as follows:

$$dz = z_x\, dx + z_y\, dy = p\, dx + q\, dy$$

To do this we need the values of p and q. As we integrate out the curve we also have to keep track of the values of the gradient. We can calculate the increments in p and q by:

$$dp = p_x\, dx + p_y\, dy \text{ and } dq = q_x\, dx + q_y\, dy$$

At first, we appear to be getting into more difficulty, since now we need to know $p_x$, $p_y = q_x$ and $q_y$. In order to determine these unknowns we will differentiate the basic equation $a(p,q) = b(x,y)$ with respect to x and y:

$$a_p\, p_x + a_q\, q_x = b_x \text{ and } a_p\, p_y + a_q\, q_y = b_y$$

While these equations contain the right unknowns, there are only two equations, not enough to solve for three unknowns. Note, however, that we do not really need the individual values! We are only after the linear combinations $(p_x dx + p_y dy)$ and $(q_x dx + q_y dy)$.

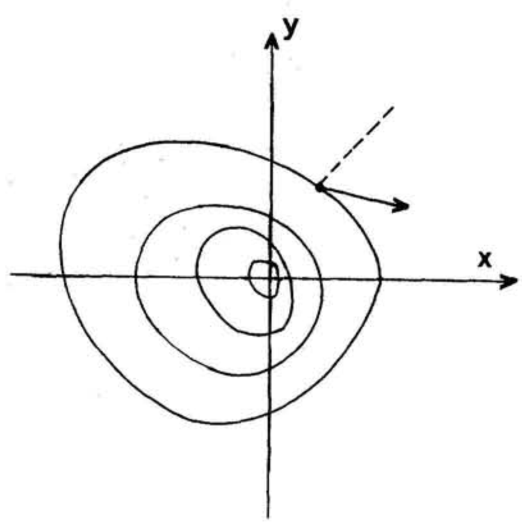
We have to choose the direction of the small step $(dx,dy)$ properly to allow the determination of these quantities. There is only one such direction. Let $(dx,dy) = (a_p,a_q)ds$, then $(dp,dq) = (b_x,b_y)ds$. This is the solution we were after. Summarizing, we have five ordinary differential equations:

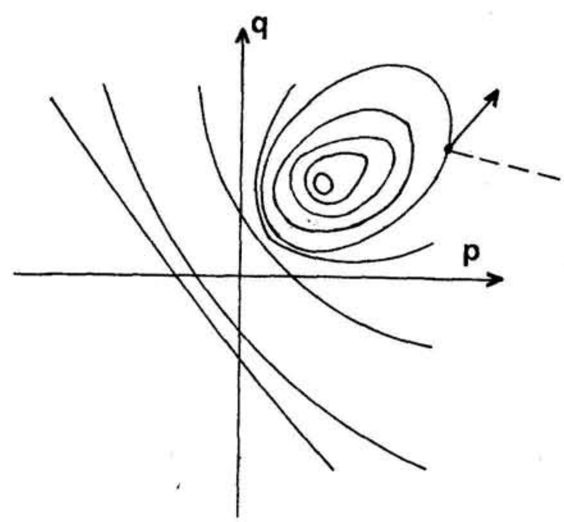
$$\dot{x} = a_p,\ \dot{y} = a_q,\ \dot{z} = pa_p + qa_p,\ \dot{p} = b_x, \text{ and } \dot{q} = b_y$$

Here the dot denotes differentiation with respect to s, a parameter that varies along the solution curve.

INTERPRETATION IN TERMS OF THE GRADIENT-SPACE:

As we solve along a particular characteristic curve on the object, we simultaneously trace out a base characteristic in the image and a curve in gradient-space. At each point in the solution we will know which point in the image and which point in the gradient-space the surface element under consideration corresponds to. The intensity in the real image and in the gradient-space image must, of course, be the same. The paths in the two spaces are related in a peculiar manner. The step we take in the image will be perpendicular to the contour in gradient-space and the step we take in gradient-space will be perpendicular to the intensity contour in the real image.



## GENERALIZATION TO NEAR SOURCE AND NEAR VIEWER:

The last solution method, while correct for arbitrary reflectivity functions, still assumes orthographic projection and a distant source. This is a good approximation for many practical cases. In order to take into account the effects of the nearness of the source and the viewer, we have to discard the gradient-space diagram, since it is based on the assumption of constant phase angle. The problem can still be tackled by algebraic manipulation, much as the last solution. It turns out that one is really trying to solve a first order non-linear partial differential equation in two independent variables. The well-known solution involves converting this equation into five ordinary differential equations, quite like the ones we obtained in the last section [4].

## ACKNOWLEDGMENTS:

I wish to thank Kathy E. Van Sant for an early version of the numerical solution for the truncated-plane mutual-illumination problem and Karen Prendergast for preparing the drawings.

REFERENCES:


1. D. A. Huffman. "Impossible Objects as Nonsense Sentences," in Machine Intelligence 6, R. Meltzer and D. Michie (editors), Edinburgh University Press, pp. 295-323.

2. A. K. Mackworth. "Interpreting Pictures of Polyhedral Scenes," in Artificial Intelligence, Vol. 4, 1973, pp. 121-137.

3. D. A. Huffman. "Curvature and Creases: A Primer on Paper," Proceedings of Conference on Computer Graphics, Pattern Recognition and Data Structures, May 1975, pp. 360-370.

4. B. K. P. Horn. "Obtaining Shape from Shading Information," in The Psychology of Machine Vision, P. H. Winston (editor), McGraw-Hill, 1975, pp. 115-155.

5. J. Van Diggelen. "A Photometric Investigation of the Slopes and Heights of the Ranges of Hills in the Maria of the Moon," Bull. Astron. Inst., Netherlands, Vol. 11, No. 423, 1951, pp. 283-289.

6. T. Rindfleisch. "Photometric Method for Lunar Topography," Photogrammetric Eng., Vol. 32, No. 2, 1966, pp. 262-276.

7. A. Herskovits and T. O. Binford, "On Boundary Detection, MIT Artificial Intelligence Memo 183, July 1970, pp. 19, 55, 56.

8. Bui Tuong Phong, "Illumination for Computer Generated Pictures," CACM, Vol. 18, No. 6, June 1975.

9. D. Hilbert and S. Cohn-Vossen. Geometry and the Imagination, Chelsea Publishers, New York, 1952.

10. F. D. Hildebrand. Methods of Applied Mathematics, Prentice-Hall, New Jersey, 1952, pp. 222-294.

# MATHEMATICAL DETAILS

## DUAL-SPACE:

One approach to gradient-space is to consider it as a projection of dual space [1,2,9]. Dual-space is a three-dimensional entity obtained by mapping planes into points and points into planes. A point of course can be specified as a vector (x,y,z). A plane also can be defined in terms of a vector (a,b,c). The plane consists of points which satisfy the equation:

$$(x,y,z)\cdot(a,b,c) = 1 \text{ or } ax + by + cz = 1$$

It is clear that a plane in one space can be mapped into a point in the other and that, conversely, a point can be mapped into a plane. These operations are reversible, that is, if we start with a plane, find the corresponding point in dual-space, we can map this point back into the original plane.

What about lines? Lines can be thought of either as the intersection of two planes or as connections between two points. Thus, the dual of a line, considered to be formed by the intersection of two planes, can be construed to be the line connecting the two points in dual-space that correspond to these two planes. A line also can be associated with the family of all planes passing through it -- its dual will be the line formed by mapping all of these planes into points.

What does the corner of a polyhedron correspond to in dual-space? First of all, a corner is a point, so it must map into a plane. Secondly, it lies in each of the planes intersecting to form the corner, so its dual must contain all of the points corresponding to these planes. The dual of a corner is the plane defined by the points corresponding to the planes that intersect to form the corner. The edges of the object meeting at the corner map into lines connecting these points.

The object-space is not directly accessible to us, since we have only a projection of it, the image-space. We cannot expect to arrive at the results in dual space simply and directly -- but it turns out that a very useful projection of dual-space exists.

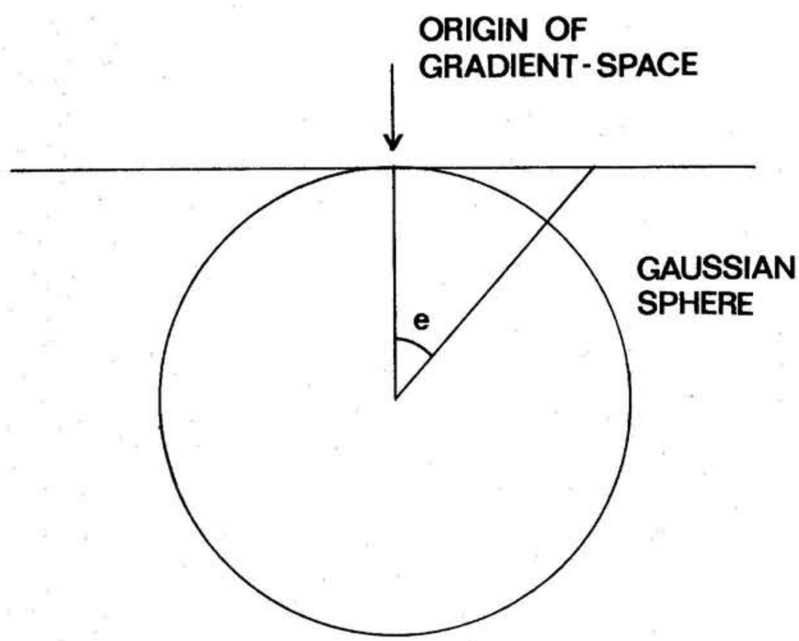
Given a point $(a,b,c)$ in dual-space, one can define its projection into gradient-space as $(-a/c,-b/c)$. This is a perspective projection. How is this related to the original object-space? Let a plane in object-space be defined as $ax + by + cz = 1$. This can also be written:

$$z = (-a/c)x + (-b/c)y + (1/c)$$

It is clear now why $(p,q) = (-a/c,-b/c)$ is called the gradient of the plane. In fact, $p = z_x$ and $q = z_y$, the first partial derivatives of z with respect to x and y respectively.

## THE GAUSSIAN SPHERE:

Another convenient way to talk about directions is by way of a unit sphere surrounding the point in question [3,9]. Points on the sphere then define specific directions. This representation is very convenient for some purposes since some useful invariants exist on the surface of this sphere which are lost in projection. Huffman uses this to advantage in analyzing developable surfaces [3]. For our purposes, however, a planar representation is more convenient. Gradient-space is simply a projection of the Gaussian sphere, with the center of the sphere acting as the center of projection and the projection being constructed onto a plane tangent to the sphere.



THE GRADIENT-LINE IS PERPENDICULAR TO THE IMAGE-LINE:

Consider two planes defined by the equations:

$$a_1x + b_1y + c_1z = d_1 \text{ and } a_2x + b_2y + c_2z = d_2$$

These planes have normals $(a_1,b_1,c_1)$ and $(a_2,b_2,c_2)$ respectively. The planes intersect in a line. The direction of this line can be found by taking the cross-products of the two normals. This follows from the fact that the line of intersection certainly has to be in both planes and, hence, perpendicular to both normals. The cross-product turns out to be $(b_1c_2-b_2c_1,a_2c_1-a_1c_2,a_1b_2-a_2b_1)$.

The image-line is the orthogonal projection of the line of intersection. Its direction is simply $(b_1c_2-b_2c_1,a_2c_1-a_1c_2)$.

The two planes map into the points $(-a_1/c_1,-b_1/c_1)$ and $(-a_2/c_2,-b_2/c_2)$ in gradient-space. The line connecting these two points is the gradient-line. Its direction can be found by subtraction to be $(a_2/c_2-a_1/c_1,b_2/c_2-b_1/c_1)$.

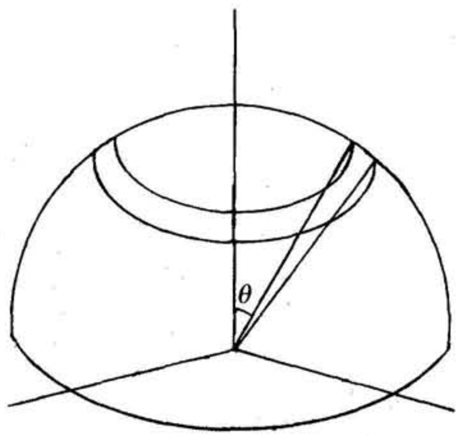
In order to establish that the gradient-line so defined is perpendicular to the image-line, we have to show that the dot-products of their respective directions is zero.

$$(b_1c_2-b_2c_1,a_2c_1-a_1c_2)\cdot(a_2c_1-a_1c_2,b_2c_1-b_1c_2)/(c_1c_2) = 0$$

The two lines are thus perpendicular. The same result can be developed using only geometric reasoning.

## INTEGRAL OF $\cos^n(\theta)$ OVER A HEMISPHERE:

To provide for the correct scaling of the specular component of reflected light we need the integral of $\cos^n(\theta)$ over the hemisphere $0 \leq \theta \leq \pi/2$.




The area of the strip on the surface of the hemisphere is $2\pi R^2 \sin(\theta) d\theta$. Integrating, we get:

$$\int_0^{\pi/2} 2\pi R^2 \sin(\theta)\cos^n(\theta)d\theta$$

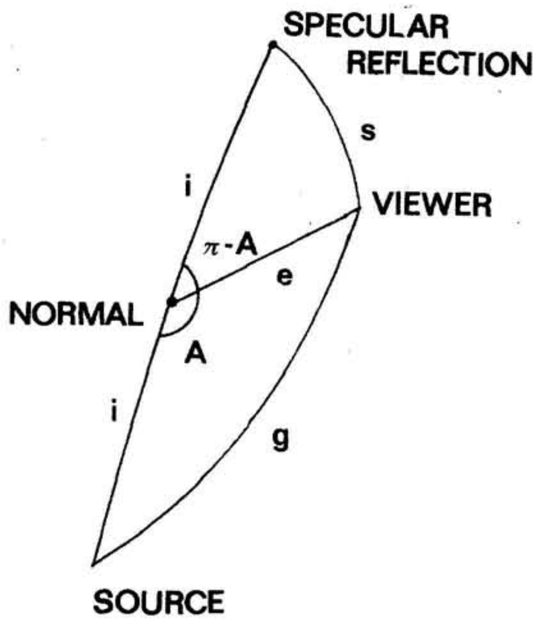
$$2\pi R^2 \int_0^{\pi/2} \cos^n(\theta)\sin(\theta)d\theta$$

$$2\pi R^2 \left[-\frac{\cos^{n+1}(\theta)}{n+1}\right]_0^{\pi/2} = 2\pi R^2/(n+1)$$

This is $1/(n + 1)$ of the surface area of the hemisphere.

THE OFF-SPECULARITY ANGLE:

For surfaces with a specular component of reflectivity one needs to know the angle between the reflected ray and the line of sight [8]. This angle can be found by simple application of some results of spherical geometry.



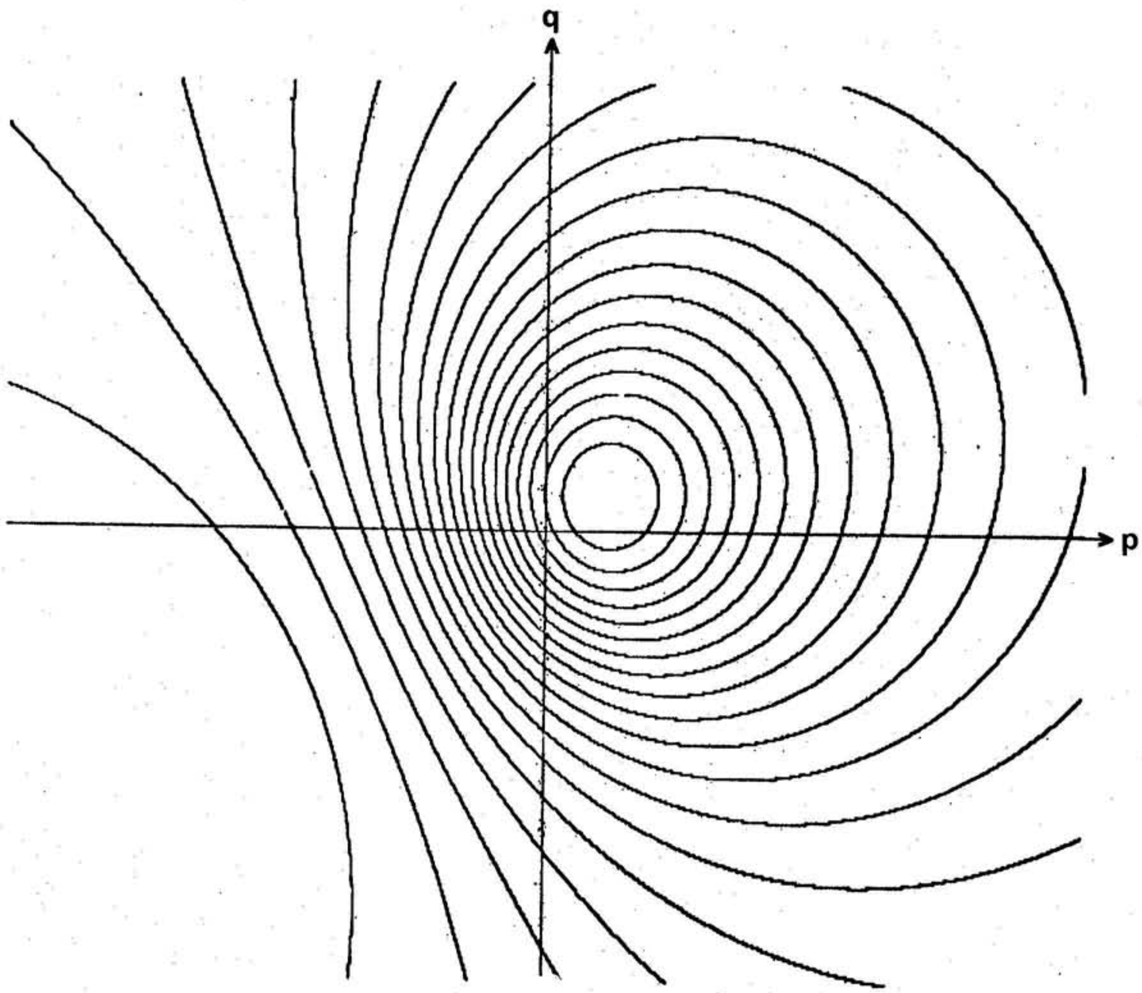

Here A is called the azimuth angle.
We are given i, e, and g, and have to find the angle s.

$$\cos(s) = \cos(i)\ \cos(e) + \sin(i)\ \sin(e)\ \cos(\pi - A)$$
$$\cos(g) = \cos(i)\ \cos(e) + \sin(i)\ \sin(e)\ \cos(A)$$

Clearly, $\cos(s) = 2\cos(i)\ \cos(e) - \cos(g) = 2IE - G$

If $2IE - G = k$ and $c = G/(k + G)$, then, $(p - p_s c)^2 + (q - q_s c)^2 = \dfrac{1 - k^2}{(k + G)^2}$.

So, the contours of constant (2 E - G) are circles. This also follows from the circle-preserving property of stereographic projection.




Contours of (2IE-G). The contour intervals are .1 units.

GRADIENT-SPACE IMAGE FOR A METALLIC SURFACE:

For specular reflection we must have two constraints satisfied; the incident angle has to equal the emittance angle, and the incident ray, the emitted ray and the surface normal have to be coplanar.

$$\text{If } i = e, \text{ then } I = E \text{ and so } (1 + p_s p + q_s q)EG = E$$

And so,

$$(1 + p_s p + q_s q) = \sqrt{1 + p_s^2 + q_s^2}$$

Next we must have $(p,q,-1)$, $(p_s,q_s,-1)$, and $(0,0,-1)$ co-planar. That is, the dot-product of any one with the cross-product of the other two must equal zero. Expressed another way, we must have the volume of the parallelepiped defined by the three vectors equal zero. Or, finally:

$$\det \begin{vmatrix} p_s & q_s & -1 \\ p & q & -1 \\ 0 & 0 & -1 \end{vmatrix} = 0, \text{ that is, } p_s q - q_s p = 0.$$

The same result could be arrived at in a more round-about fashion by requiring that $i + e = g$, and then expanding $\cos(i + e) = \cos(g)$. We now have two linear equations in p and q:

$$p_s p + q_s q = \sqrt{1 + p_s^2 + q_s^2} - 1$$

$$q_s p - p_s q = 0$$

Solving for p and q:

$$p = (\sqrt{1 + p_s^2 + q_s^2} \ -1) \ p_s/(p_s^2 + q_s^2) = p_s \ \frac{G}{1 + G} \quad = \cos\theta\sqrt{(1 - G)/(1 + G)}$$

$$q = (\sqrt{1 + p_s^2 + q_s^2} \ -1) \ q_s/(p_s^2 + q_s^2) = q_s \ \frac{G}{1 + G} \quad = \sin\theta\sqrt{(1 - G)/(1 + G)}$$

This related to the half-angle formula:

$$\tan(\tfrac{1}{2}\alpha) = \sqrt{(1 - \cos\alpha)/(1 + \cos\alpha)}$$

STEREOGRAPHIC PROJECTION:

The gradient-space image for a metallic object is a stereo-graphic projection of the surround of the object. That is, the sphere of possible direction as seen by the object is mapped onto a place, with the center of projection at one pole of the sphere and the plane tangent at the other pole. The mapping is conformal; that is, angles are preserved. Circles on the sphere are mapped mapped into circles on the plane. The following illustrations from pages 248, 252, and 253 of Hilbert & Cohn-Vossen [9] will illustrate:

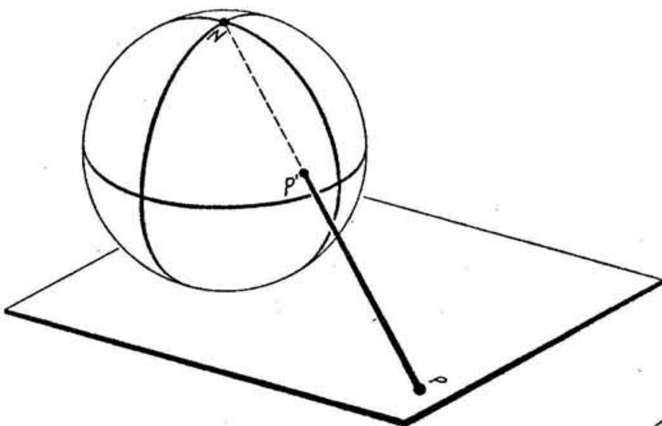



FIG. 239

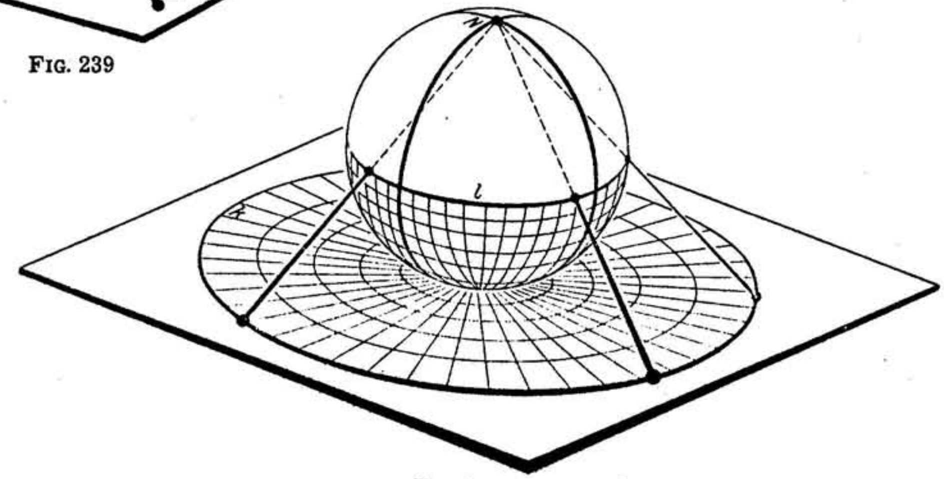



FIG. 244a

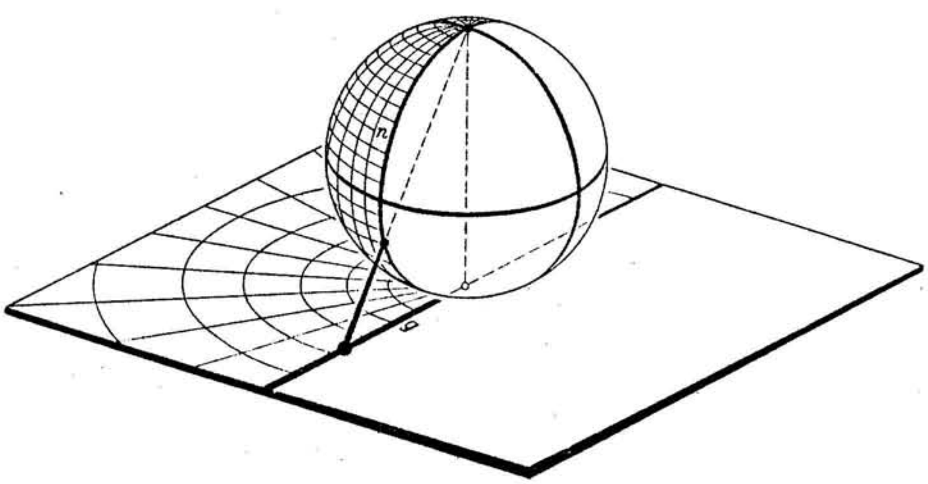



FIG. 244b

A USEFUL DISCRIMINANT:

The incident, emittance, and phase angle form a spherical triangle and have to satisfy certain constraints -- that is, we cannot arbitrarily choose i, e, and g. The sum of any two has to exceed the third. This is analogous to a similar result for the sides of planar triangles. It is easy to see that only one of the three constraints can fail at any one time [4]. Suppose it is the following:

$$i + e < g, \text{ then } \cos(i + e) > \cos(g)$$

since cosine is monotonically decreasing in the range 0 to $\pi$. Expanding, one gets:

$$\cos(i)\cos(e) - \cos(g) > \sin(i)\sin(e)$$

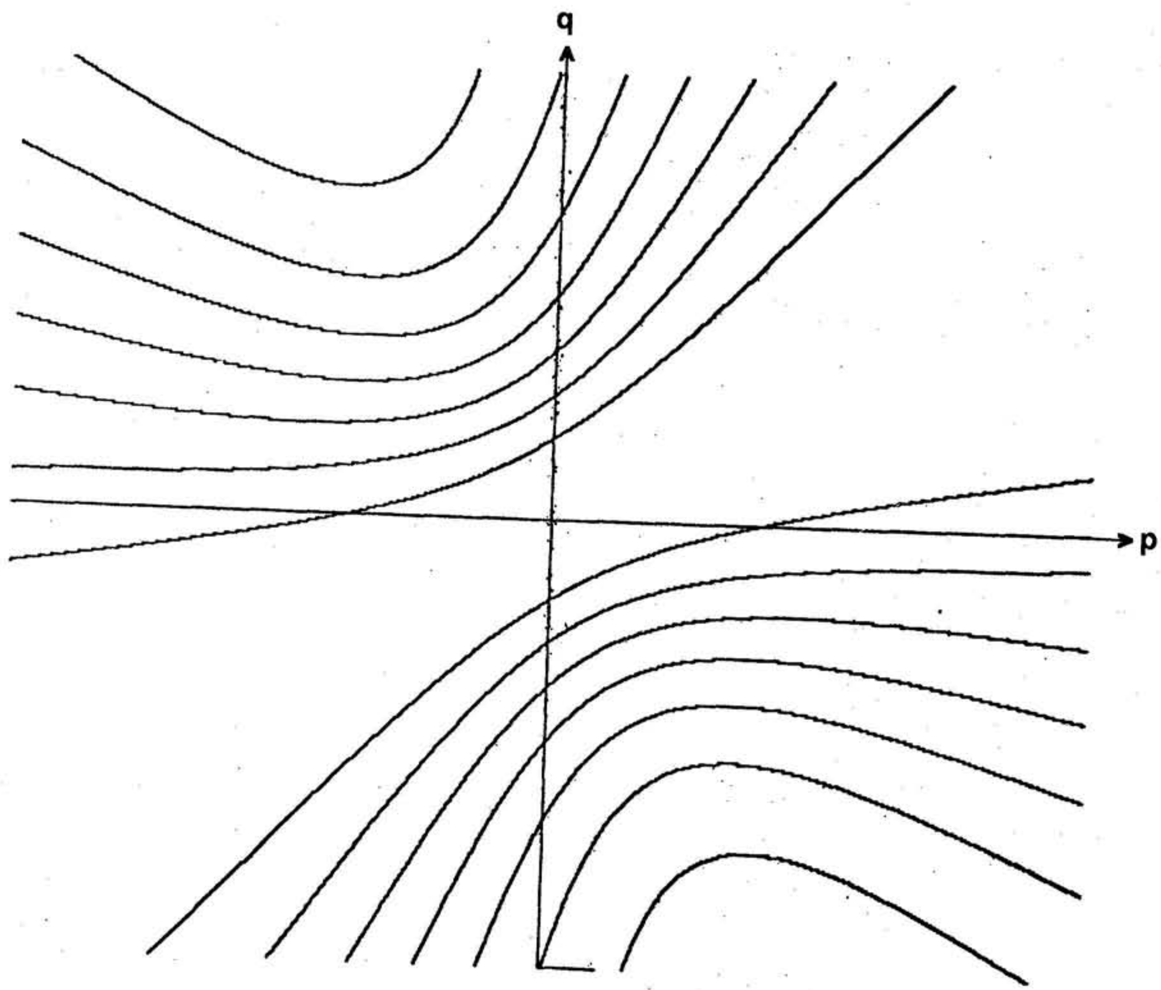
The righthand side is positive, so we can square both sides, hence:

$$(IE - G)^2 > (1 - I^2)(1 - E^2) \text{ or } 1 + 2IEG - (I^2 + E^2 + G^2) < 0.$$

The symmetry of this expression suggests that we would have obtained the same result if we had picked either of the other two constraints. In fact, it is easy to show that if i, e, and g can form a spherical triangle, then

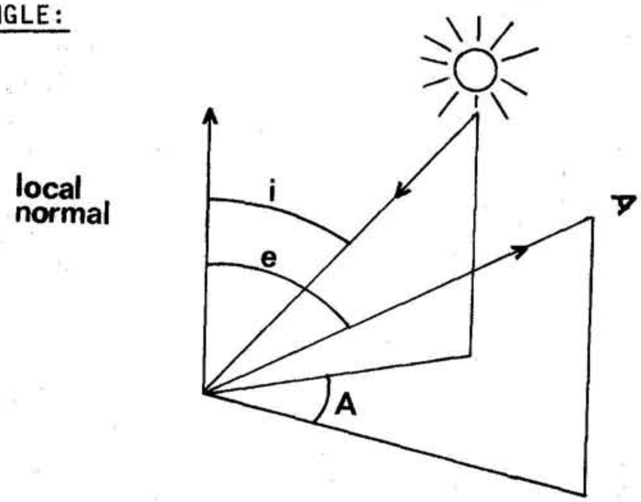
$$1 + 2IEG - (I^2 + E^2 + G^2) \geq 0$$

and that this expression is less than zero otherwise.




Contours of the discriminant $1 + 2IEG - (I^2 + E^2 + G^2)$. The contour intervals are .05 units.

THE AZIMUTH ANGLE:




A useful quantity for some manipulators is the azimuth angle A, between the projections of the incident and emitted rays onto the object's surface. Applying a result of spherical trigonometry, we find

$$\cos(g) = \cos(i)\cos(e) + \sin(i)\sin(e)\cos(A)$$

So,

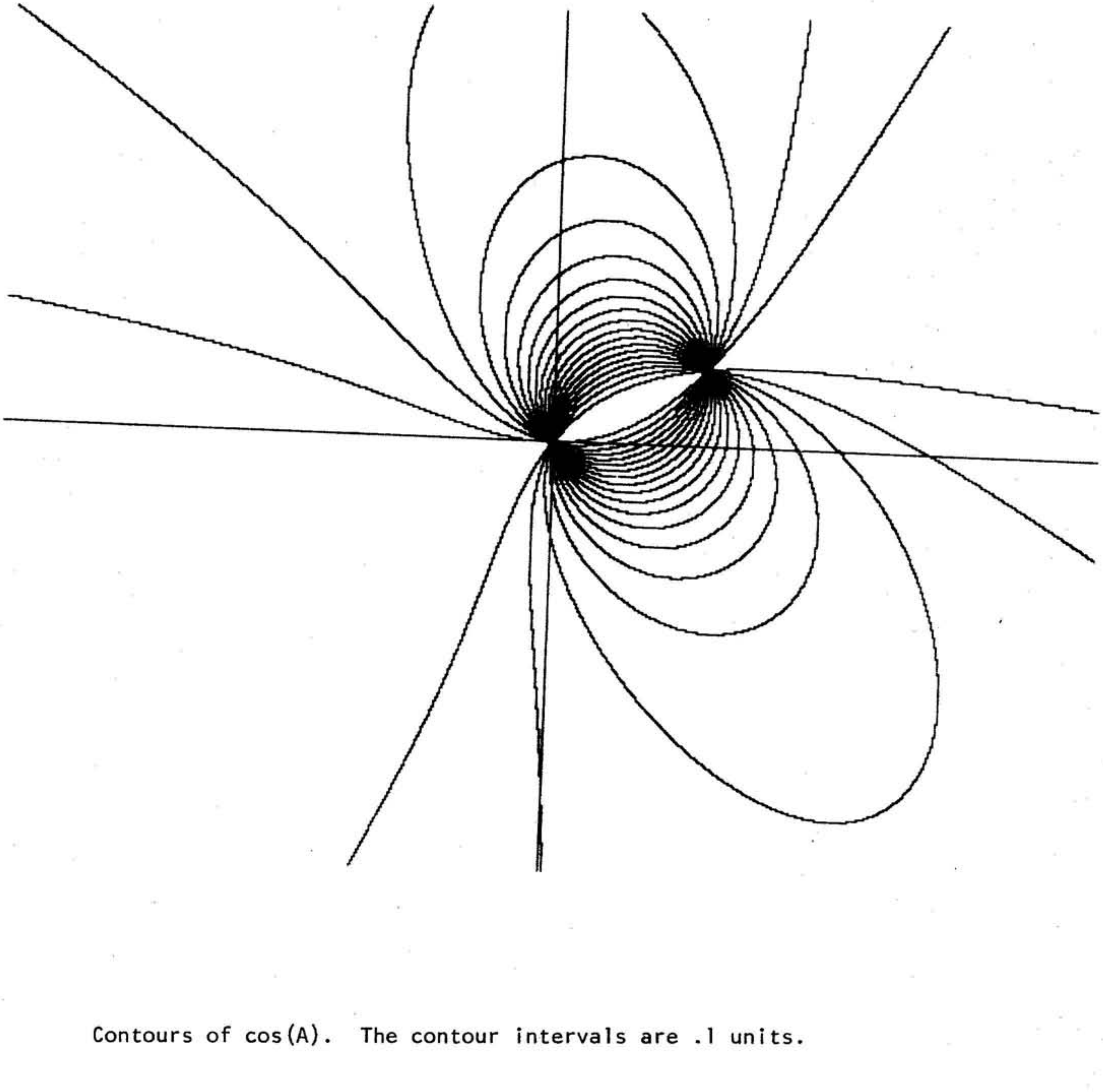
$$\cos(A) = \frac{G - IE}{\sqrt{1 - I^2}\,\sqrt{1 - E^2}}$$

Now obviously, $\cos(A) \leq 1$

Expanding, $(G - IE)^2 \leq (1 - I^2)(1 - E^2)$

And so again,

$$1 + 2\,EG - (I^2 + E^2 + G^2) \geq 0$$



Contours of cos(A). The contour intervals are .1 units.

EXPANDING THE DISCRIMINANT:

We would like to express the discriminant in terms of $p$, $q$, $p_s$, and $q_s$. We will need the following:

$$G = 1/\sqrt{1 + p_s^2 + q_s^2}, \text{ and } E = 1/\sqrt{1 + p^2 + q^2}$$

Let $\quad X = (1 + p_s p + q_s q)$, then $I = XEG$

Then,

$$1 + 2IEG - (I^2 + E^2 + G^2) = 1 + 2XE^2G^2 - (X^2E^2G^2 + E^2 + G^2)$$

$$= -E^2G^2 + 2XE^2G^2 - X^2E^2G^2 + 1 + E^2G^2 - E^2 - G^2$$

$$= -E^2G^2(1 - X)^2 + (1 - E^2)(1 - G^2)$$

$$= [(1/E^2 - 1)(1/G^2 - 1) - (1 - X)^2]E^2G^2$$

$$= [(p^2 + q^2)(p_s^2 + q_s^2) - (p_s p + q_s q)^2]\ E^2G^2$$

$$= (q_s p - p_s q)^2\ E^2G^2 \qquad (!)$$

It is immediately apparent that the discriminant is positive for all points in gradient-space, as it should be. But what is more exciting is that we have an equation that is linear in $p$ and $q$ and thus helpful if we are going to try to obtain $p$ and $q$, given $I$, $E$, and $G$:

$$q_s p - p_s q = \pm\sqrt{1 + 2IEG - (I^2 + E^2 + G^2)}\ /EG$$

FINDING p AND q, GIVEN I, E, AND G:

We now have two linear equations in p and q, one from the expression for I, the other from the expansion of the discriminant:

$$p_s p + q_s q = (I/E - G)/G$$

$$q_s p - p_s q = \pm(\Delta/E)/G$$

where $\Delta^2 = 1 + 2IEG - (I^2 + E^2 + G^2)$. Solving for p and q we get:

$$p = \frac{(I/E - G)/Gp_s/(p_s^2 + q_s^2) \pm (\Delta/E)/Gq_s}{(p_s^2 + q_s^2)}$$

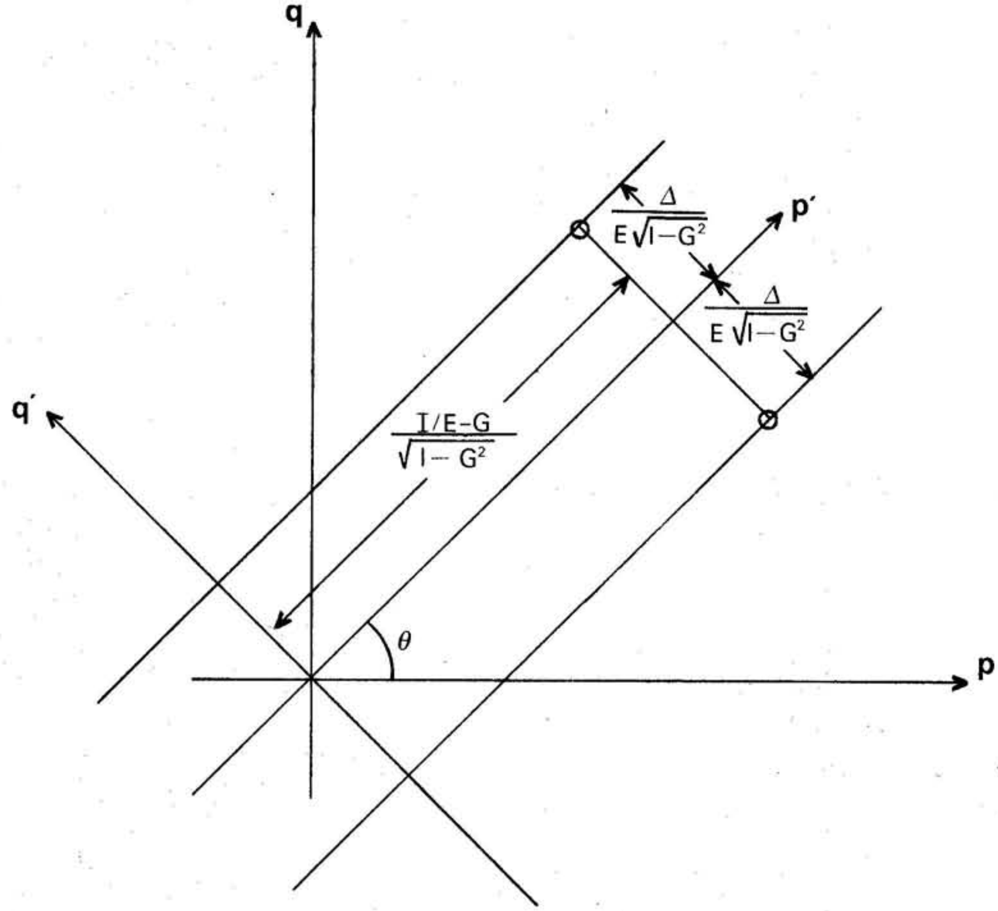
$$q = \frac{(I/E - G)/Gq_s/(p_s^2 + q_s^2) \pm (\Delta/E)/Gp_s}{(p_s^2 + q_s^2)}$$

If we let $\dfrac{p_s}{\sqrt{p_s^2 + q_s^2}} = \cos(\theta)$, and $\dfrac{q_s}{\sqrt{p_s^2 + q_s^2}} = \sin(\theta)$,

$p' = \dfrac{(I/E - G)}{\sqrt{1 - G^2}}$, and $q' = \pm \dfrac{(\Delta/E)}{\sqrt{1 - G^2}}$, then

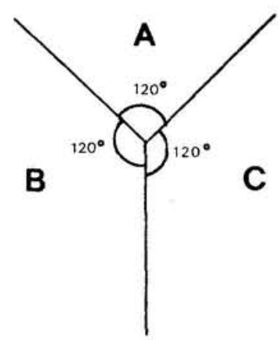
$$p = p' \cos(\theta) - q' \sin(\theta)$$

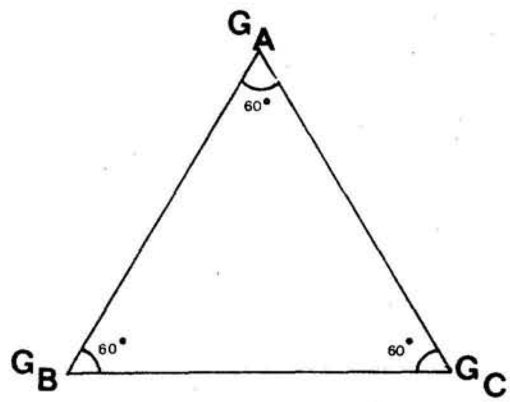
$$q = p' \sin(\theta) + q' \cos(\theta)$$


q
p′
q′
p
$\theta$
$\frac{\Delta}{E\sqrt{1-G^2}}$
$\frac{\Delta}{E\sqrt{1-G^2}}$
$\frac{I/E-G}{\sqrt{1-G^2}}$

DETERMINATION OF ORIENTATION OF PLANES:

An example will illustrate how image intensity information can augment Mackworth's gradient space scheme for interpreting polyhedral scenes. We are given a trihedral corner projected into the image as follows:




The corresponding gradient-space diagram with position and scale as yet undetermined is on the right. Now we are told that the (normalized) image intensities are .79, .30, and .86 for the regions A, B, and C respectively. We thus have six constraints on the position of the three points in gradient space. Given that $(p_s, q_s) = (0.7, 0.3)$, and $\phi(I,E,G) = I$, we can develop the following equations:

$$q_B = q_C$$

$$(q_A - q_B) = +\sqrt{3}\,(p_A - p_B)$$

$$(q_A - q_C) = -\sqrt{3}\,(p_A - p_C)$$

$$(1 + .7p_A + .3q_A) = \sqrt{1 + p_A^2 + q_A^2}\ \frac{.79}{.80}$$

$$(1 + .7p_B + .3q_B) = \sqrt{1 + p_B^2 + q_B^2}\ \frac{.30}{.80}$$

$$(1 + .7p_C + .3q_C) = \sqrt{1 + p_C^2 + q_C^2}\ \frac{.86}{.80}$$

Where we used the fact that $G \approx .80$. Squaring the second set of three equations, we obtain second-order polynomials. This simply reflects the fact that the points are constrained to lie on certain conic sections. Using an iterative modified Newton-Raphson method, one quickly converges to a solution as follows:

$$(p_A, q_A) \approx (0, .70)$$

$$(p_B, q_B) \approx (-.61, -.35)$$

$$(p_C, q_C) \approx (+.61, -.35)$$

The polynomials are actually simple enough that they might be solved directly using appropriate symbol manipulation algorithms.

The following questions are left as an exercise for the reader:

1. Is there another solution?
2. Are there solutions for which the three edges are concave?
3. Are the points $G_B$, $G_C$ as precisely determined in gradient-space as the point $G_A$?

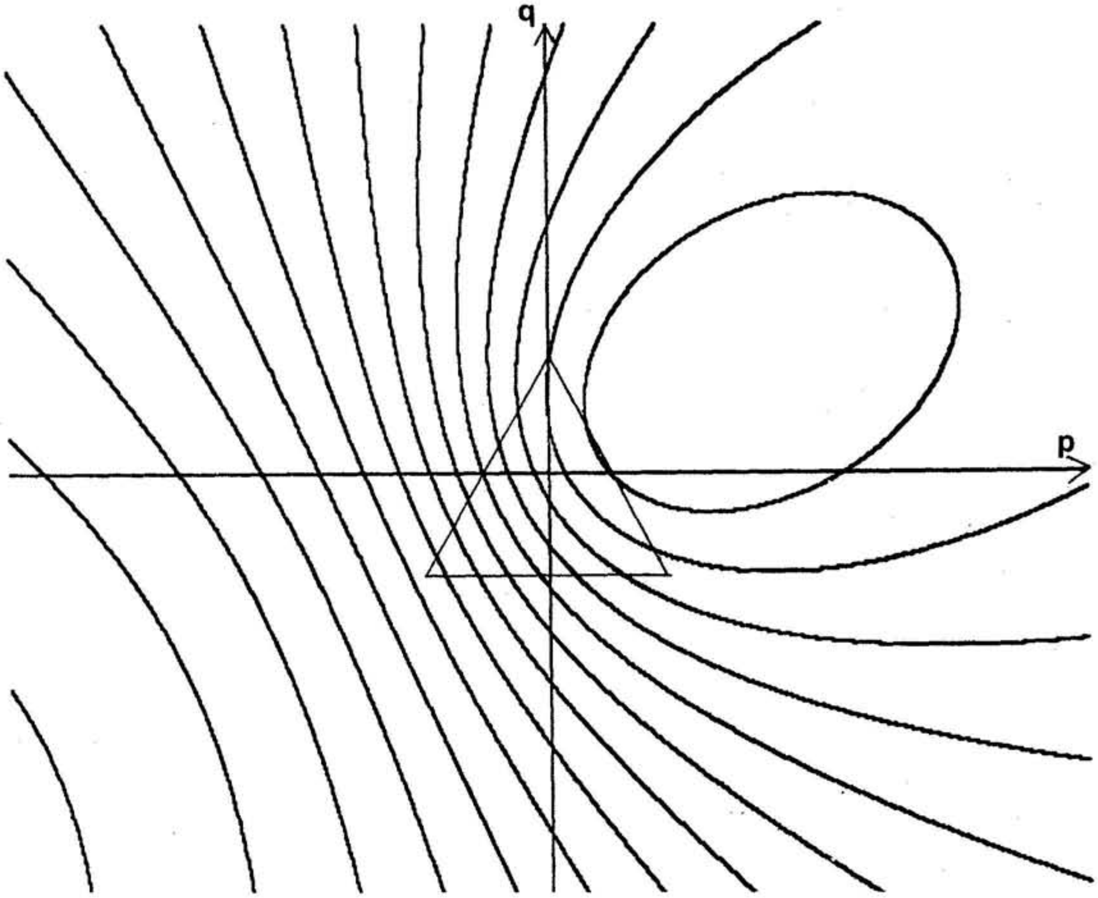
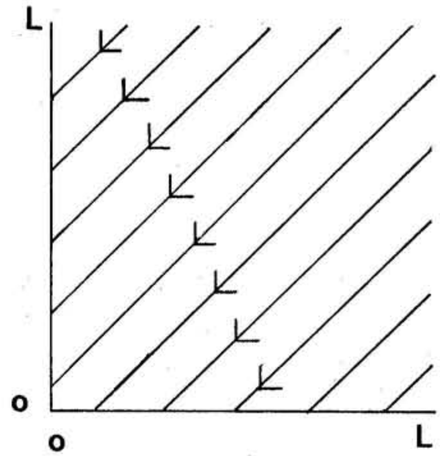
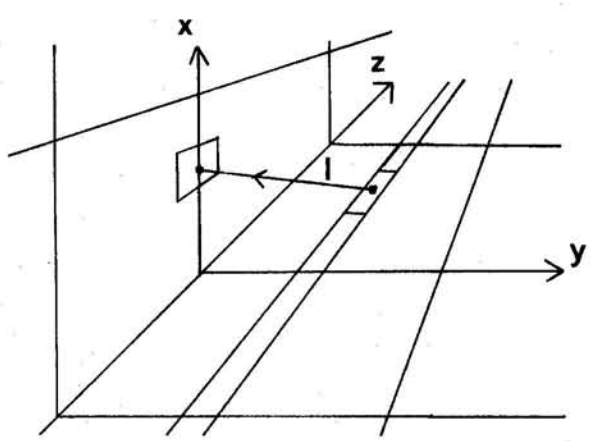



Illustration of the constraints on the three gradient-space points $G_A$, $G_B$, and $G_C$. This is the solution to the problem of determining the orientation of the three faces meeting at the corner.

MUTUAL ILLUMINATION -- PLANES TRUNCATED IN ONE DIRECTION:




In order to get a feel for the mutual illumination problem it helps to study a simple case first. Consider two planes joined at right angles, infinite in the direction of the line of their intersection, and both of length L in the direction away from their intersection. Let the incident light come from a distant source and in a direction $\pi/4$ with respect to the planes. This last condition and the equal lenth of the sides provide the symmetry necessary to ensure that the intensity distribution on the two planes is equal. Next we will assume that the surfaces are lambertian, with reflectivity r.



Now let us calculate the total light flux received by a surface element a distance x away from the corner. First consider the contribution due to a surface element on the other plane a distance y from the corner and separated by a distance z along the direction of the line of intersection. Let the luminous emittance vary as $L(y)$. Then this contribution will be, for lambertian surfaces,

$$(r/\pi)[\cos(i)\cos(e)]/l^2\ L(y)dydz \qquad \text{(flux/unit area)}$$

Here l is the distance between the two points, e is the angle of emittance at the emitting surface element and i is the angle of incidence at the receiving surface element.

$$\cos(e) = x/l,\ \cos(i) = y/l,\ \text{and}\ l^2 = x^2 + y^2 + z^2$$

So the contribution due to the patch (dy by dz) is then:

$$L(y)dy(r/\pi)(xy)/l^4\ dz$$

Integrating with respect to z, one obtains:

$$L(y)dy(r/\pi)xy \int 1/(x^2 + y^2 + z^2)^2\ dz$$

Now, $$\int 1/(a^2 + s^2)^2\ ds = (1/a^3)(\pi/2)$$

So we get: $$L(y)dy\ \tfrac{1}{2} r\ xy/(x^2 + y^2)^{3/2}$$

INTEGRAL EQUATION:

Finally integrating with respect to y and adding in the direct contribution:

$$L(x) = E/\sqrt{2} + \tfrac{1}{2} r \int_0^L \frac{xy}{(x^2 + y^2)^{3/2}} L(y)dy$$

So here we have an implicit equation for L(x) called an integral equation. Before we try to solve it, notice that the parameters E and L can be eliminated. For example, if L(x) is a solution for incoming light flux E, then aL(x) will be a solution if the light-flux is changed to aE. That is, everything just gets brighter in proportion, if we increase the incident flux. Without a loss of generality, we can set $E/\sqrt{2} = 1$.

Next, let $x' = x/L$ an $y' = y/L$, then we find

$$L(x') = 1 + 1/2r \int_0^1 \frac{x'y'}{(x'^2 + y'^2)^{3/2}} L(y')dy'$$

So, we can, without loss of generality, also let L = 1. So we will try to solve:

$$L(x) = 1 + \tfrac{1}{2} r \int_0^1 \frac{xy}{(x^2 + y^2)^{3/2}} L(y)dy$$

This is a Fredholm integral equation of the second kind [10]. Such equations usually occur as solutions to ordinary differential equations with given end conditions. The kernel is,

$$K(x,y) = \frac{xy}{(x^2 + y^2)^{3/2}}$$

The kernel is symmetric, non-separable, and worst of all, not bounded. There are a number of techniques for solving such equations with symmetric kernels, but most work only for bounded kernels, or, if one can calculate the iterated kernel.

ITERATIVE SOLUTION:

One method is iteration [10]. Suppose we start with $L_o(x) = 0$. Substituting this in the righthand side of the equation we arrive at the next approximation: $L_1(x) = 1$. Using this we integrate again and find:

$$L_2(x) = 1 + \tfrac{1}{2} r\left(1 - \frac{x}{\sqrt{1 + x^2}}\right)$$

The next step leads to:

$$L_3(x) = 1 + \tfrac{1}{2} r(1 + \tfrac{1}{2} r)\left(1 - \frac{x}{\sqrt{1 + x^2}}\right) - (\tfrac{1}{2} r)^2 \, x\int_o^1 \frac{1}{(x^2 + y^2)^{3/2}} \, \frac{y^2}{(1 + y^2)^{1/2}} \, dy$$

This last term turns out to be some messy difference of elliptic integrals and so we abandon further iteration. A few things of note emerge, however. First of all, we have a useful first approximation in $L_2(x)$ or the first few terms in $L_3(x)$. This approximation is particularly good for small r, since the remaining omitted terms are in $r^2$ and higher orders.

Secondly, the leading term will clearly become on further iteration:

$$1 + \tfrac{1}{2} r + (\tfrac{1}{2} r)^2 + (\tfrac{1}{2} r)^2 + \ldots \quad 1/(1 - \tfrac{1}{2} r)$$

And so, $L(0) = 1/(1 - \tfrac{1}{2} r)$, not too surprisingly. Next one can say something about the convergence of this iterative process -- it will converge for r less than 2, and diverge for r greater than or equal to 2. Obviously, we care only about values for r between zero and one, so we expect a solution will always exist.

INTEGRATION BY PARTS:

Further useful results can be obtained by integrating by parts:

$$L(x) = 1 + 1\tfrac{1}{2}\, r \left[L(0) - L(1) \frac{x}{\sqrt{1 + x^2}} + \int_0^1 L'(y) \frac{x}{\sqrt{x^2 + y^2}}\, dy\right]$$

Clearly $L(0) = 1 + 1\tfrac{1}{2} r\ L(0)$ and so, once again, we see that $L(0) = 1/(1 - 1\tfrac{1}{2} r)$. This result depends on the fact that the integral is zero for $x = 0$, which can be shown by applying L'Hospital's rule to the resulting indeterminate form. By a tedious method of little interest here, one can arrive at another approximation:

$$L(x) = 1 + \frac{\tfrac{1}{2}r}{1 - \tfrac{1}{2}r} \left[1 - x^{(1 - \tfrac{1}{2}r)}\right]$$

This approximation is particularly good near the origin and can be "tuned" by multiplying the factor containing x by a number smaller than one. The form of this result shows that $L(x)$ will have a cusp at the origin for r greater than or equal to one.

NUMERICAL SOLUTION:

It is becoming increasingly obvious that an analytical solution is not around the corner, so it is time to turn to numerical methods. There are again various possible avenues, the most obvious being iteration -- since we already know some good first approximations we can speed up the convergence. The only difficulty is the singularity in the kernel for $x = y = 0$. Dividing the range of integration evenly produces quite poor results particularly near the origin for large values of $r$. Dividing the range more finely near the origin and ignoring the first few values near there is the obvious solution. Choosing as end-points of the intervals the points $(i/n)^2$ works quite well. Here $n$ is the total number of segments in the interval from 0 to 1. The mid-points of the intervals are used when evaluating the kernel.
The resulting solutions for $n = 256$ were presented graphically earlier.