Astron. Astrophys. 259, 318–332 (1992)

ASTRONOMY AND ASTROPHYSICS

Interpretation of lightcurves of atmosphereless bodies

I. General theory and new inversion schemes

M. Kaasalainen [1,2], L. Lamberg [3], K. Lumme [1], and E. Bowell [4]

[1] University of Helsinki, Observatory and Astrophysics Laboratory, Tähtitorninmäki, SF-00130 Helsinki, Finland
[2] University of Oxford, Department of Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, England
[3] University of Helsinki, Department of Mathematics, Hallituskatu 15, SF-00100 Helsinki, Finland
[4] Lowell Observatory, Mars Hill Road, 1400 West, Flagstaff, AZ 86001, USA

Received October 10, 1991; accepted January 9, 1992

Abstract. We present theoretical methods of lightcurve inversion that can be used in photomorphography, i.e., determination of the three-dimensional shape and/or the light-scattering behaviour of the surface of a body from disk-integrated photometry. These methods can be applied to atmosphereless bodies – mainly asteroids and planetary satellites – in the solar system. With no loss of generality, the objects of our study are referred to as asteroids in this paper.

The inversion comprises three steps. First, a function (or functions) containing information about both the shape and the albedo variegation of an asteroid is determined. This step is feasible provided the surface is strictly convex and provided a sufficient number of lightcurves are available at different observing geometries and at nonzero phase angles. Also, the functional form of the surface light-scattering law must be known and it must be of a suitable type. This inversion problem is mathematically ill-posed; i.e., small errors in the data may have large effects on the results. Also, the number and range of the observing geometries have a significant effect on the inversion.

In the second step, separate expressions for the inverse of the Gaussian curvature and the albedo distribution are derived from the information obtained in the first step. This is possible if the functional form of the scattering law as a function of albedo is known and it is of a suitable form (a necessary but not sufficient requirement for this is nonlinearity in albedo). In the third step, the non-trivial problem of determining the radius vector of the surface from the Gaussian curvature is solved by using iterative optimization procedures developed by us.

Key words: asteroids – photometry – analytical methods

1. Introduction

The lightcurve of a rotating asteroid is determined by the asteroid's shape, light-scattering behaviour, and the observing geometry. The goal of photometric morphography (or photomorphography) of asteroids is to express these dependencies in such a way that a solution for the inverse problem can be sought.

Send offprint requests to: M. Kaasalainen (University of Oxford)

The first analysis of the inverse problem was offered by Russell (1906). As he pointed out, it is, of course, "quite impossible to determine the shape of the asteroid" if we consider the problem in its most general form, i.e., with no restrictions on the parameters. Russell primarily studied the opposition situation with geometric scattering (the observed brightness of a geometrically scattering surface is directly proportional to the projected area of the surface in the direction of the observer): if only zero solar phase angle lightcurves are available, an infinite variety of shapes or surface albedo distributions will account for the observed lightcurves even with restrictions on the parameters. Russell's conclusions were thus rather pessimistic, but this was because of the very limited choice of the observing geometry and because of the somewhat restrictive mathematical methods Russell used.

However, as we shall show, if the problem is studied with nonzero phase angles (or at opposition, provided the scattering is nongeometric) and certain reasonable assumptions are made, there are methods for obtaining a solution.

After Russell, not much work was done on the subject. One recent approach is that of Ostro & Connelly (1984, 1988). This method produces a two-dimensional "convex profile" of an asteroid and does not employ lightcurve measurements at many different observing geometries, whereas our aim here is to use as many geometries as possible to obtain a directly comprehensible model of an asteroid. Some studies of the direct problem have been made in order to investigate mainly qualitatively the effects that various shape and surface properties have on lightcurves; interesting relationships are presented e.g. in Cellino et al. (1989) and Karttunen & Bowell (1989).

A very important matter in all inverse problems is to study the effect of the quantity and the quality of the observational data on the solution. Our purpose is to pay close attention to this often rather improperly treated matter in our studies.

In Sect. 2, we pose the problem in exact terms and discuss some assumptions that have to be used. We formulate the direct problem by expressing mathematically how the observed brightness of an asteroid depends on the light-scattering behaviour and the shape of the asteroid and on the observing geometry. In Sect. 3, we discuss some general properties of the inverse problem before formulating the inversion methods at opposition and at general observing geometries in Sects. 4 and 5. In Sect. 6, we introduce the essential principles of our methods for obtaining the

shape of a surface from the inverse of its Gaussian curvature. In Sect. 7, we present two numerical simulations to illustrate and discuss the practical implementation of our techniques. In the next paper, we study more closely the role of the a priori assumptions used in inversion, and we investigate the effects of various properties of observational data on the inversion in order to give a realistic estimate of the goodness of the solution in different cases; we shall also apply our techniques to real lightcurve data. In Sect. 8, we present our conclusions and discuss some future work and applications.

For understanding the main points of this treatment, Sects. 3, 5.2, and 6 are not necessary and can be skipped on a first reading.

2. Posing the problem

First, we assume that the observing geometry is known; i.e., that the positions of the Sun and Earth relative to the asteroid (these are easily calculated) and the asteroid's pole position or a good first approximation to it are known. We also assume relaxed rotation – that the asteroid is not precessing – and that the absolute rotational phase of the asteroid can be computed for any moment. Thus the sidereal rotation period of the asteroid must be known. Solving for the pole position is a nontrivial problem: we refer mainly to the method of Lumme et al. (1990) because of its model independence and its "separability" from rotation period and shape/albedo problems. There are, of course, several other methods (for a summary, see Magnusson et al. 1989). Accurate determination of the rotation period is also a difficult task (see, e.g., Karttunen & Muinonen 1991), but it is vital for photomorphography. This is because photomorphographic methods make use of a large number of lightcurves obtained at different observing geometries; yet asteroids may typically make hundreds of revolutions between apparitions, thus inducing large errors in the absolute rotational phase if the sidereal rotation period is not very accurately known.

In theory, the pole position/period/shape/albedo problems should be solved simultaneously to obtain the highest accuracy. As we shall show, the shape/albedo part of the inverse problem may, using reasonable assumptions, be considered linear (the integrated brightness of an asteroid is a linear functional of functions describing the surface features). On the other hand, the rotation period/pole position part of the problem is nonlinear. In practice, this gives rise to a simultaneous solution of the joint inverse problem. First, approximation values for the nonlinear parameters (pole direction/rotation period) are determined and then the inversion methods for the linear part, described in this paper, are applied using a series of small deviations from the previously determined values of the nonlinear part, so that the solution that gives the best fit in this series can be chosen.

In studying the problem of lightcurve inversion many different geometrical configurations have to be used; thus we use several different coordinate systems, described in Appendix A. Coordinate systems $C1$–$C3$ describe the Sun-Earth-asteroid geometry and coordinate system $C4$ describes the surface of the asteroid (in parametrizing a surface, two variables are needed). In this connection, we state the most important restriction imposed on the properties of the asteroid in this paper: that the surface of the asteroid is assumed to be strictly convex (mathematically, this is also called an ovaloid surface); in other words, convex without planar sections. On such a surface there is one and only one point corresponding to a given surface normal direction. This assumption is necessary, because only in this case can the integration limits of the integral describing the asteroid's brightness be written as constants, which is essential for the inversion.

2.1. Light-scattering law

Firstly, we need an expression giving the observed light flux $F_{\rm obs}$ from a surface element of area $\mathrm{d}\sigma$ on the asteroid when the flux incident on the element is $F_{\rm in}$. This is written as

$$F_{\rm obs} = \frac{F_{\rm in}}{r^2} S(\mu, \mu_0, \alpha; \boldsymbol{P})\,\mathrm{d}\sigma, \tag{2.1}$$

where r is the distance between the Earth and the element and $S(\mu, \mu_0, \alpha; \boldsymbol{P})$ is the scattering law for the element. The cosines of the angles between the surface normal and the directions of the Earth and Sun are denoted, respectively, by μ and μ_0, and α is the solar phase angle. The physical parameters in the scattering law are represented by the set $\boldsymbol{P}$. These parameters may depend on the location on the surface. In practice, the most important physical parameter – at least in the lightcurve inversion – is the single scattering albedo ϖ_0, because in many scattering laws the effects it produces are indistinguishable from those of body shape. However, as we shall show, if the scattering law is of a suitable form, these effects are separable.

In what follows, we shall for the sake of clarity use the albedo ϖ_0 to represent any physical parameter in the scattering law. The addition of any other physical parameters to the scattering law is straightforward in the formalism we use.

In its most general form the scattering law should be written as $S(\mu, \mu_0, \alpha;\ \varpi_0;\ u, v)$, the coordinates (u, v) parametrizing the surface, so that the locational dependence on the surface is not necessarily coupled to the physical parameters. However, we shall show that to uniquely determine a function describing the shape, this coupling must at least partially obtain, so in practice we may write $S = S(\mu, \mu_0, \alpha; \varpi_0(u, v))$.

The scattering law may also be written as

$$S(\mu, \mu_0, \alpha; \varpi_0) = \mu\mu_0 R(\mu, \mu_0, \alpha; \varpi_0), \tag{2.2}$$

where the effect of the reflection coefficient R is separated from that of the attitude of the surface element with respect to the Sun and the Earth. Since $F_{\rm in}$ and r are always known, they can be eliminated from the lightcurve (or, equivalently, the lightcurve can be normalized to some values of $F_{\rm in}$ and r). Thus the observational quantity is the scattering cross-section or (normalized) brightness L of the surface element:

$$L = S(\mu, \mu_0, \alpha; \varpi_0)\,\mathrm{d}\sigma. \tag{2.3}$$

The construction of a physically consistent scattering law is a demanding task. The problems involved include, among other things, modelling the scattering properties of single regolith particles and using the theory of radiative transfer in connection with multiple scattering. In practice, certain approximations must be made. In photomorphography, the general functional form of the scattering law is assumed known. This, of course, is one possible source of error in the inversion. Examples of some applicable scattering laws (or hypotheses) are the Lommel-Seeliger law (which gives geometric scattering at opposition), Lambert's law, Hapke's law and the Lumme-Bowell law (see, e.g., Bowell et al. 1989; and Lumme et al. 1980).

The scattering law describes the microscopic structure of the surface statistically. Small craters may also be included in this description in a statistical manner. The convexity assumption means convexity on a global scale. As some simulations (Kart-

tunen 1989; Karttunen & Bowell 1989) have shown, very local "non-statistical" deviations from convexity, such as larger craters, often make no significant contributions to lightcurves and are thus not real obstacles for inversion under the convexity assumption.

2.2. *Integrated brightness*

Consider a surface in any system of parametric coordinates that define it uniquely, the coordinates being u, v. Denote

$$\boldsymbol{J}(u,v) = \frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}, \tag{2.4}$$

where $\boldsymbol{r}(u,v)$ is the radius vector of the surface. The vector $\boldsymbol{J}(u,v)$ is parallel to the surface normal $\boldsymbol{n}(u,v)$ at the point (u,v).

In the previous section, we defined the observed (normalized) brightness of a surface element. The integrated brightness consists then of the summed brightnesses of all the illuminated and observable elements:

$$L(\boldsymbol{E},\boldsymbol{E}_0,\boldsymbol{r},S) = \iint_{S_+} S(\mu,\mu_0,\alpha;\varpi_0)\,|\boldsymbol{J}(u,v)|\,\mathrm{d}u\,\mathrm{d}v \tag{2.5}$$

integrated over the part S_+ of the surface for which $\mu, \mu_0 \geqq 0$.

In the inverse problem, information on S and $\boldsymbol{r}$ is sought using the integral Eq. (2.5), when some sets of values for L are known (from lightcurves obtained at various positions). In evaluating the integral, we can use two geometrically consistent coordinate systems for the variables u, v:

1. Spherical coordinates in the asteroid's reference frame, the z-axis being the rotation axis. In the inverse problem, an expression for $\boldsymbol{J}$ [and thus for $\boldsymbol{n}(u,v)$] in this system enables the shape to be easily constructed. However, the integration limits become very complex and approximations deriving from, e.g., a triaxial ellipsoid have to be used, a procedure that is not suitable for irregular shapes. Therefore, this coordinate system is not suitable for the inversion, although it can be used in the direct calculation of lightcurves when the shape is known.

2. A spherical coordinate system for the surface normal in which a point on the surface is defined by the polar spherical coordinates (ϑ, ψ) of the direction of the surface normal. Since the surface is strictly convex, this mapping (also called the Gaussian image) from the surface onto a unit sphere is unique. In this case we need to know the function

$$G(\vartheta,\psi) = \frac{|\boldsymbol{J}(\vartheta,\psi)|}{\sin\vartheta}. \tag{2.6}$$

Another expression for this is $1/K(\vartheta,\psi)$, where K is the Gaussian curvature. Using normal coordinates, the quantity $G\sin\vartheta$ gives the area of a surface element at the point (ϑ,ψ), and thus the function G is called the Gaussian surface density. In shape/pattern recognition in computer science G is also called the extended Gaussian image (Horn 1984), which term we prefer to use in the discrete case concerning the areas and normal directions of the faces of a polyhedron. This coordinate system enables us to obtain an expression for $G(\vartheta,\psi)$ relatively easily, because the integration limits can always be made constant by rotating the coordinate system. The Gaussian surface density defines a strictly convex surface unambiguously, but the problem of constructing the shape from $G(\vartheta,\psi)$ is nontrivial. This parametrization allows construction of a function uniquely describing the surface assuming only that the surface is convex.

Using normal coordinates (ϑ,ψ), the integral (2.5) can now be written as

$$L(\boldsymbol{E},\boldsymbol{E}_0,\boldsymbol{r},S) = \iint_{S_+} S(\mu,\mu_0,\alpha;\varpi_0)\,G(\vartheta,\psi)\,\sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\psi. \tag{2.7}$$

The integration region S_+ is that part of the surface where the angle between the line of sight and the surface outward normal and the angle between the direction of the Sun and the outward normal are both between the limits $[-\frac{\pi}{2},\frac{\pi}{2}]$. Thus, in the Gaussian image, the integration zone is the intersection of two hemispheres whose polar directions lie in the directions of the Sun and the Earth.

3. General considerations

3.1. *The inverse problem*

The integral equation (2.7) is a Fredholm equation of the first kind, and serves as a typical example of an ill-posed inverse problem. The ill-posed nature of the problem arises from the smoothing behaviour of the integral: variations in body shape (or G) are mapped onto relatively smaller variations in the lightcurve L. From this it follows that uncertainties in the lightcurve data are projected onto larger variations in the evaluated G. More precisely, the integral (2.7) as an operator is compact, so the inverse operation is unbounded. The source function, i.e., the function obtained by inversion from a function describing the data, can be discontinuous even though the data function is continuous. This is a mathematical situation often encountered in problems of inverting remotely sensed data. Such problems are common in natural sciences, especially in astronomy and geophysics (for some astronomical examples, see Craig & Brown 1986).

The ill-posed nature of the problem becomes very manifest when the integral (2.7) is written discretely in linear form

$$\boldsymbol{L} = A\boldsymbol{g}, \tag{3.1}$$

where the components of the vector $\boldsymbol{L}$ are observed brightnesses, the components of the vector $\boldsymbol{g}$ are discrete average values of the Gaussian surface density on small surface patches on the Gaussian unit sphere, and the matrix A corresponds to the integration operation. The chosen inversion solution could be the one for which the norm $\|\boldsymbol{L} - A\boldsymbol{g}\|$ describing the misfit is smallest (a least-squares fit when the norm is Euclidean). There are, however, lots of quite different solutions for which the misfit value is well below the observational uncertainty, and thus there is no way of determining the correct solution without using some a priori constraints and/or information. The very essence of solving ill-posed problems lies in the use of a priori assumptions. To stabilize the solutions in such problems, various regularization schemes using physical a priori knowledge (or assumptions) can be used. Since regularization methods are very commonly used, they are described at length in many sources (a good, compact introduction is Craig & Brown 1986). In what follows, we shall pay attention to the stability of our inversion schemes. Where a stabilizing method is needed, our choice of a suitable regularization method is the so-called statistical inversion (e.g. Menke 1984), based on Bayesian estimation theory. This scheme is very advantageous in the sense that it gives the inversion solution in the form of a so-called a posteriori distribution; i.e., it provides an estimate of the goodness of the solution. We shall describe this method in our next paper concerning lightcurve inversion.

It is worthwhile to look into the form of the scattering law a little more closely here. As stated earlier, there may be locational dependence (explicit dependence on the variables parametrizing the surface) in the scattering law as well, e.g., in the form of albedo

variegation. If this dependence is complicated, so that the scattering law is a nonlinear functional of location-dependent functions, the inversion is very difficult to perform. If the scattering law is a linear functional of location-dependent functions, a feasible inversion procedure can be developed. Expressed in another way, the integral equation (2.7) can now be written

$$\boldsymbol{L} = M\boldsymbol{p}, \tag{3.2}$$

where the matrix M is formed from the location-independent part of the scattering law and the vector $\boldsymbol{p}$ may contain several vectors representing functions that depend on the location on the surface. The Gaussian surface density is included in the vector $\boldsymbol{p}$. As we shall show below, it is possible to separate the Gaussian surface density uniquely from the other location-dependent functions in the cases where the locational dependence in the scattering law, rather than being independent of the physical parameters, is at least partially coupled.

3.2. Representation of a function as a Laplace series

In theory, the solution of the inverse problem may be sought using the discrete form (3.1). In this case, however, hundreds or thousands of unknown variables must be used. Of course, instability of the solution is thereby increased, necessitating the use of suitable regularization schemes where, e.g., the expected smoothness of the solution is used as a constraint.

Another way of expressing a function to be determined is to represent it as a functional series, the coefficients of which are now the unknown variables. In theory, such a series is infinite, but the number of variables can be greatly reduced if the series is supposed to converge rapidly so that it can be truncated. Using suitable functions, this can often be expected in physically realistic situations. In fact, this assumption is in itself a rather powerful regularization method, corresponding to the aforementioned smoothness constraint. We shall show that its application can lead to a stable solution of the inverse problem if we have at our disposal a large number of lightcurves measured at various observing geometries.

The use of the coefficients of a functional series produces one practical difficulty in this case. The function obtained in inversion (e.g., the Gaussian surface density or the albedo distribution) should normally be everywhere positive by definition; however, there usually are no simple ways of describing this restriction in the function coefficient space in order to incorporate it in the inversion formalism. Thus the positivity of the solution must be checked separately after the coefficients have been obtained. If the solution turns out to be substantially negative somewhere, one must use more or less intuition in re-regularizing the solution to be positive or in concluding whether the surface is substantially nonconvex.

Before investigating the form of a suitable functional series more closely, let us consider the integral (2.7) generally. Because the integration zone is simply an intersection of two hemispheres, the integration limits can be transformed into constants using a suitable rotation of the coordinate system. These rotations are described in detail in the sections below concerning the opposition situation and the general situation. We parametrize a rotation using the Euler angles described in Appendix B. After rotating a coordinate system, the functions in the integrand of (2.7) must be transformed accordingly, in such a way that their values will remain unchanged. Thus, when a vector $\boldsymbol{x}$ is transformed as

$$\boldsymbol{x} \to \boldsymbol{x}'; \ \boldsymbol{x}' = R\boldsymbol{x}, \tag{3.3}$$

where R is the operator (matrix) performing the rotation, the value of a scalar function f in a point in space must remain unchanged. Thus

$$f'(\boldsymbol{x}') = f(\boldsymbol{x}), \tag{3.4}$$

where $f'(\boldsymbol{x}')$ is the function in the new coordinate system. This form can always be obtained from any arbitrary function f by using

$$f'(\boldsymbol{x}) = f(R^{-1}\boldsymbol{x}), \tag{3.5}$$

fulfilling the condition (3.4). In inversion, however, use of (3.5) usually leads to very complicated functions in the integrand of (2.7), so analytical integration becomes impossible. In this case, the problem can be posed as in (3.1) or (3.2), the matrix operator now being computed by numerical integration and the unknowns being the coefficients of a functional series.

The aforementioned inversion scheme is quite implementable and requires no specific properties of the chosen functional series. However, it is advantageous to consider the choice of the functional series further. Since we are operating on the unit sphere of the Gaussian image, a spherical harmonics series (also called a Laplace series) is a very natural choice. Spherical harmonics have the remarkable property of transforming as a finite-sized linear combination of themselves under rotations of the coordinate system, as shown in Appendix B. This transformation thus separates the parameters describing the rotation from the variables of a function, which is a very useful property in inversion, allowing further analytical inspection of the integral (2.7).

4. The opposition situation revisited

In this section, we represent Russell's (1906) main results using a more modern mathematical formalism. This formalism is then extended in the next section to include observing geometries at non-zero phase angles.

The integral (2.7) can now be written as

$$L(\boldsymbol{E}, \boldsymbol{r}) = \iint\limits_{\mu \geqq 0} S(\mu; \varpi_0)\, G(\vartheta, \psi) \sin\vartheta \, \mathrm{d}\vartheta \, \mathrm{d}\psi . \tag{4.1}$$

In the coordinate system $C1$, the direction of the Earth (and the Sun) is expressed using coordinates (θ, φ). Let us now perform a rotation such that the new z-axis will lie in the direction of the Earth. The directions of the new x- and y-axes are irrelevant here. Using Euler angles (Appendix B), this rotation is $R(0, \theta, \varphi)$. Now $\mu = \mu_0$ and after rotation

$$\mu = \cos\vartheta . \tag{4.2}$$

The scattering law is denoted by $S(\mu; \varpi_0)$. If there is dependence on the location on the surface in the scattering law in the simple form

$$S(\mu; \varpi_0(\vartheta, \psi)) = S_0(\mu)\, P(\vartheta, \psi) \tag{4.3}$$

(e.g., albedo in Russell's text), we write the product of the Gaussian surface density $G(\vartheta, \psi)$ and the function P as a Laplace series

$$G(\vartheta, \psi)\, P(\vartheta, \psi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} b_{lm}\, Y_l^m(\vartheta, \psi) . \tag{4.4}$$

If P is a constant, then b_{lm} are the coefficients of the Laplace series describing the Gaussian surface density. Since the series is real-valued, only the coefficients $b_{lm}, m \geqq 0$ are of interest, the other

coefficients being completely determined by these. The new form $S_0'(\vartheta,\psi)$ [or here $S_0'(\vartheta)$] of the function S_0 in the rotated system is obtained using (4.2).

If P_R is an operator transforming a function in any arbitrary rotation $R(\gamma,\beta,\alpha)$, we have

$$P_R(\gamma,\beta,\alpha)\ Y_l^m(\vartheta,\psi)=\sum_{m'=-l}^{l} Y_l^{m'}(\vartheta,\psi)\ D^{(l)}_{m'm}(\gamma,\beta,\alpha) \tag{4.5}$$

(for a description of $D^{(l)}_{m'm}$ see Appendix B), so the integral (4.1) becomes

$$\begin{aligned} L(\theta,\varphi)&=\int_0^{2\pi}\int_0^{\pi/2} S_0'(\vartheta)\sum_{lm} b_{lm}\sum_{m'} D^{(l)}_{m'm}(0,\theta,\varphi)\\ &\quad\cdot Y_l^{m'}(\vartheta,\psi)\sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\psi\\ &=2\pi\sum_{lm} b_{lm}\int_0^{\pi/2} Y_l^0(\vartheta,\psi)\ S_0'(\vartheta)\\ &\quad\cdot D^l_{0m}(0,\theta,\varphi)\sin\vartheta\,\mathrm{d}\vartheta\,. \end{aligned} \tag{4.6}$$

Because $Y_l^0(\vartheta,\psi)=P_l(\cos\vartheta)$, we obtain, using (B9),

$$L(\theta,\varphi)=2\pi\sum_{lm} k_l b_{lm}\ Y_l^m(\theta,\varphi)\,, \tag{4.7}$$

where

$$k_l=\int_0^{\pi/2} S_0'(\vartheta)\ P_l(\cos\vartheta)\sin\vartheta\ \mathrm{d}\vartheta\,. \tag{4.8}$$

Since scattering laws in opposition are typically proportional to terms of the form μ^n, we derive for this case a general expression for the coefficients k_l.

Assume the opposition form S_0 of a scattering law to be $S_0=\mu^n$, n an integer. Equation (4.8) can now be written

$$k_l^{(n)}=\int_0^1 x^n\ P_l(x)\ \mathrm{d}x\,. \tag{4.9}$$

Using the recursion relations

$$(2l+1)\ P_l(x)=P'_{l+1}(x)-P'_{l-1}(x) \tag{4.10}$$

and

$$(l+1)\ P_{l+1}(x)=(2l+1)\ xP_l(x)-lP_{l-1}(x) \tag{4.11}$$

and integrating by parts, one arrives at a recursion relation

$$k_l^{(n)}=\frac{n}{l+n+1}\,k_{l-1}^{(n-1)} \tag{4.12}$$

for values $n>0, l>0$. The starting values are

$$k_0^{(n)}=\frac{1}{n+1} \tag{4.13}$$

and for odd l

$$k_l^{(0)}=(-1)^{(l-1)/2}\,\frac{(l-2)!!}{(l+1)!!} \tag{4.14}$$

while for even $l>0$ $k_l^{(0)}$ vanishes, as can easily be calculated. If n is not an integer, $k_l^{(n)}$ can be computed from the series representation

$$k_l^{(n)}=\sum_{i=0}^{[l/2]}(-1)^i\,\frac{(2l-2i)!!}{2^l i!\,(l-i)!\,(l-2i)!}\,\frac{1}{l-2i+n+1}\,. \tag{4.15}$$

As can be seen from the relation (4.12), e.g., for geometric scattering $[S(\mu)=a\mu]$ the value of k_l is zero for each odd $l>1$, whereas for Lambert scattering $[S(\mu)=b\mu^2]$ the value of k_l is zero for each even $l>2$. Thus, in these cases, the observed lightcurves depend on, and thus contain information on, only a part of the unknown coefficients b_{lm}. In other words, the inverse solution is unique only if the scattering law consists of terms containing both even and odd exponents for μ.

If there are sufficiently many lightcurves obtained at different aspect angles θ (most preferably in equally spaced intervals, thus making use of the discrete Fourier orthogonality), a two-dimensional Fourier series representing the data can be constructed. This series can readily be transformed into a Laplace series

$$L(\theta,\varphi)=\sum_{lm} h_{lm}\ Y_l^m(\theta,\varphi) \tag{4.16}$$

using the orthogonality relation (B6) of spherical harmonics. As was noted in connection with (4.8), odd harmonics $l>1$ are absent in the lightcurve in the case of geometric scattering. The unknown coefficients b_{lm} can be directly obtained from Eq. (4.7):

$$b_{lm}=\frac{h_{lm}}{2\pi k_l}\,. \tag{4.17}$$

This, of course, requires that the integrals k_l be non-zero. Thus when the scattering is geometric the unknown Laplace series (4.4) cannot be uniquely determined. In any case, in order to determine the Gaussian surface density, the scattering law has to be known explicitly.

The values of the coefficients k_l at increasing values of l provide a clear manifestation of the principal source of instability in the inversion. As can be seen from Eqs. (4.12)–(4.14), the values of k_l decrease rapidly with increasing l, which means that the larger the degree l of a coefficient b_{lm}, the less it contributes to the integrated brightness: thus, information of the high-degree components of the shape is easily drowned in the noise. In other words, Eq. (4.17) shows that in inversion the errors in b_{lm}, induced by the errors in the observed coefficients h_{lm}, are rapidly amplified with increasing l, thus causing instability if the series is not truncated early.

If the scattering law is of a more complicated form, such as

$$S(\mu;\varpi_0(\vartheta,\psi))=P(\vartheta,\psi)\sum_a a_n S_n(\mu)\,, \tag{4.18}$$

where the constants a_n are unknown, we have

$$2\pi\, b_{lm}\sum_n a_n k_l^n=h_{lm}\,, \tag{4.19}$$

in which k_l^n are computed for each S_n as in (4.8). In this case there are not enough equations to solve uniquely for the ratios of the constants a_n. Therefore one cannot determine a location-dependent function uniquely using opposition observations alone, if the scattering law is constructed from more than one scattering law so that the relative contributions of such laws are unknown.

Thus the opposition observations alone do not provide very good data for inversion. From observations and theoretical models we know that for dark objects the scattering is usually very nearly geometric at opposition, which precludes unambiguous inversion. Even if scattering were not geometric and were otherwise suitable, the scattering law would have to be known exactly.

However, as Russell noted, it is possible to draw some general conclusions on the nature of the asteroid by inspecting the observational Laplace series (4.16). Because all spherical harmonics Y_l^m with odd l change sign upon inversion of space (this

corresponds to viewing from the opposite direction), we know that if there are non-zero coefficients h_{lm} in the series (4.16) when l is odd, then

– the surface is not convex: scattering may be geometric, in which case there has to be albedo variegation, or non-geometric, in which case albedo variegation is not necessary;

– the surface is strictly convex: scattering is necessarily non-geometric if there are non-zero coefficients for which $l > 1, l$ odd.

We also know that those h_{lm} for which $l = 1$ can be induced only by non-convex shape or albedo variegation. This is because the projected areas of a closed surface in opposite directions are always equal; it can then directly be seen that the coefficients of degree $l = 1$ of a Laplace series of a Gaussian surface density are always zero. It should be noted, however, that the coefficients h_{1m} are merely a positive indicator: the *absence* of these coefficients is in principle not sufficient to indicate the absence of albedo variegation, nor can one deduce much about the albedo distribution using these coefficients. Intuitionally, one might suppose that if the coefficients vanish, the light-variations are more probably caused by shape rather than by albedo distribution.

5. General observing geometry

Consider the integral (2.7) in the general observing geometry, where the direction of the Earth is given by the coordinates (θ, φ) and the direction of the Sun by (γ, α) (obliquity and phase angle) in the coordinate system $C2$. In the coordinate system $C3$ these directions are given by $(\alpha, \delta, \varepsilon, \kappa)$. Let us now perform a rotation of the coordinate system such that the new y-axis will lie in the direction of the Earth, and the Sun will lie in the xy-plane an angle α away from the y-axis measured in the positive rotation direction.

Using general relations of spherical trigonometry we obtain

$$\mu = \sin\psi \sin\vartheta \tag{5.1}$$

and

$$\mu_0 = \sin(\psi - \alpha) \sin\vartheta . \tag{5.2}$$

Let us assume the scattering law to be of the form $S(\mu, \mu_0, \alpha)$; the possible location-dependencies will be investigated later. The rotational transformation $S'(\vartheta, \psi, \alpha)$ can be written using (5.1) and (5.2). The Gaussian surface density is expressed as a Laplace series

$$G(\vartheta, \psi) = \sum_{lm} g_{lm} Y_l^m(\vartheta, \psi). \tag{5.3}$$

Now the integral (2.7) can be written in the form

$$L(\boldsymbol{E}, \boldsymbol{E}_0, \boldsymbol{r}) = \int_{\alpha}^{\pi} \int_{0}^{\pi} F_R \, G(\vartheta, \psi) \, S'(\vartheta, \psi, \alpha) \sin\vartheta \, \mathrm{d}\vartheta \, \mathrm{d}\psi , \tag{5.4}$$

where F_R represents an operator transforming a function in the rotation. As is readily seen, the rotation can be performed as two successive rotations using the coordinate system $C2$. First the rotation $R(0, \theta - \frac{\pi}{2}, \varphi)$ is performed and then the rotation $R(0, \gamma, -\frac{\pi}{2})$, so that

$$F_R = P_R\left(0, \gamma, -\frac{\pi}{2}\right) P_R\left(0, \theta - \frac{\pi}{2}, \varphi\right), \tag{5.5}$$

where P_R is an operator transforming a function in the rotation parametrized by the given Euler angles. Thus the integral (5.4) can be written as

$$L(\alpha, \gamma, \theta, \varphi) = \sum_{lm} g_{lm} \sum_{km'} d^{(l)}_{km'}(\gamma) \cdot d^{(l)}_{m'm}\left(\theta - \frac{\pi}{2}\right) \mathrm{e}^{-\mathrm{i}m'\pi/2} \, \mathrm{e}^{\mathrm{i}m\varphi} \, I_{lk}(\alpha) , \tag{5.6}$$

where

$$I_{lk}(\alpha) = \int_{\alpha}^{\pi} \int_{0}^{\pi} S'(\vartheta, \psi, \alpha) \, Y_l^k(\vartheta, \psi) \sin\vartheta \, \mathrm{d}\vartheta \, \mathrm{d}\psi , \tag{5.7}$$

and $d^{(l)}_{m'm}$ are described in Appendix B.

Using the coordinate system $C3$, the rotation is simply $R(\kappa, \varepsilon, \delta)$, so the integral (5.4) can also be written as

$$L(\alpha, \kappa, \varepsilon, \delta) = \sum_{lm} g_{lm} \sum_{m'} d^{(l)}_{m'm}(\varepsilon) \, \mathrm{e}^{\mathrm{i}m'\kappa} \, \mathrm{e}^{\mathrm{i}m\delta} \, I_{lm'}(\alpha) . \tag{5.8}$$

Equations (5.6)–(5.8) provide a manifestation of the same property that we discussed in conjunction with (4.17): the higher the degree l of a coefficient g_{lm}, the less it contributes to the total brightness. This phenomenon gives rise to possible instability in inversion, and therefore the noise in the observations largely decides the truncation point of the Laplace series obtained in inversion. Usually, the series describing the Gaussian surface density has to be truncated early. This is not necessarily a problem in obtaining an estimate for the global shape, since with "well-behaved" surfaces the series can often be expected to converge quite rapidly.

Note that if the angles α and γ are fixed, the integrated brightness can, of course, be expressed as a Laplace series in θ and φ, but in this case (5.6) is *not* such a series. This means that the method for the opposition situation, based on the orthogonality property of spherical harmonics, cannot be generalized for use in this special case. However, we shall show that a three-dimensional generalization of this method is possible when the phase angle α is fixed.

5.1. *Inversion*

There are several possible inversion methods for the data obtained at nonzero phase angles. One method is to construct sets of simultaneous linear equations in the same manner as was discussed in connection with Eqs. (3.1), (3.2), and (3.5), the unknowns being the coefficients of a Laplace series.

Another possibility is to use the Fourier series expansions of the observed lightcurves. Equations (5.6) and (5.8) are Fourier series of the angles φ or δ when all the other observational angles are taken to be constants, thus describing lightcurves in some observational geometries. We can write this in the form

$$L(\varphi) = \sum_{lm} g_{lm} k_{lm} \, \mathrm{e}^{\mathrm{i}m\varphi} , \tag{5.9}$$

where k_{lm} includes the integral (5.7) and the functions of all the observational angles other than φ (or δ). If the observed lightcurves are written as

$$L(\varphi) = \sum_{n=-N}^{N} c_n \mathrm{e}^{\mathrm{i}n\varphi} , \tag{5.10}$$

N being the largest order included, we obtain sets of simultaneous linear equations of the form

$$c_m = \sum_{l \geqq |m|}^{L} g_{lm} k_{lm} \tag{5.11}$$

for each order m, when the Laplace series is truncated at some chosen highest degree L. The stability properties of these sets of equations are as discussed earlier.

If there are enough data to form a higher-dimensional functional series (to a sufficient accuracy), the situation is somewhat different. Assume that we have been able to form from observations a three-dimensional Fourier series for a fixed phase angle in the coordinate system $C3$; i.e., we know the observed brightness in the form $L = L(\kappa, \varepsilon, \delta)$.

Looking at Eq. (5.8), we find it to be a functional series of the elements of the rotation matrices $D^{(l)}_{m'm}(\kappa, \varepsilon, \delta)$. These elements (see Appendix B) are orthogonal and they form a complete set in the space of their arguments; i.e., a function of the Euler angles can be expanded as a series of these elements. Thus the observations at a given phase angle can be written as

$$L(\kappa, \varepsilon, \delta) = \sum_{lmm'} c^l_{m'm} D^{(l)}_{m'm}(\kappa, \varepsilon, \delta) . \tag{5.12}$$

For a strictly convex surface this follows, of course, directly from (5.8) but, as mentioned above, it holds true generally as well.

Using the orthogonality relation (B14), the observed function $L(\kappa, \varepsilon, \delta)$ can be transformed into series (5.12), so the coefficients $c^l_{m'm}$ are

$$c^l_{m'm} = (N^l_{m'm})^{-1} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} D^{(l)*}_{m'm}(\kappa, \varepsilon, \delta) \cdot L(\kappa, \varepsilon, \delta) \sin\varepsilon \, \mathrm{d}\delta \, \mathrm{d}\varepsilon \, \mathrm{d}\kappa , \tag{5.13}$$

where the normalization constant $N^l_{m'm}$ is given by (B15). Comparing (5.8) with (5.12), we obtain

$$g_{lm} = \frac{c^l_{m'm}}{I_{lm'}} . \tag{5.14}$$

Thus, with one value of the phase angle α, we obtain a maximum of $2l+1$ equations for each g_{lm}. Therefore this scheme is quite different from the methods using sets of equations: in this case we obtain separate equations for each unknown coefficient instead of having to fit all the coefficients simultaneously to best describe the data. As discussed earlier, such fitting requires regularization methods. In the case of separate equations, however, the uncertainty of an unknown coefficient depends directly on the uncertainty of the corresponding observational coefficient. The instability properties related to (5.14) are as discussed in conjunction with (4.17).

If the integral $I_{lm'}$ vanishes for some indices, we have fewer than $2l+1$ equations. In the cases of Lommel-Seeliger and Lambert scattering, for example, $I_{lm'}$ vanishes for odd $l+m'$, which results from the parity properties of the associated Legendre polynomials. Also, using the definition of the associated Legendre polynomials obtained by differentiation of the Legendre polynomials, and integrating by parts, one can show that for Lommel-Seeliger scattering all the integrals $I_{31}, I_{51}, \ldots$ vanish; for Lambert scattering the same applies for the integrals $I_{40}, I_{60}, \ldots$ and $I_{42}, I_{62}, \ldots$

If an observed coefficient $c^l_{m'm}$ is nonzero even though the corresponding $I_{lm'}$ is zero, the assumed scattering law is incorrect or the surface is nonconvex. As in the opposition situation, the coefficients $c^1_{m'm}$ can be induced only by nonconvexities or by location-dependence of the scattering law.

The analysis of the general situation differs from that of the opposition situation in that it is possible to have several equations for an unknown coefficient rather than just one. These equations can now be used to separate unknown parameters in the scattering law from the Gaussian surface density. Assume the scattering law to be of the form

$$S(\mu, \mu_0, \alpha) = \sum_n a_n(\alpha) \, S_n(\mu, \mu_0, \alpha) , \tag{5.15}$$

where the functions S_n are known or assumed, and the coefficients a_n are unknown. Now Eq. (5.14) is written as

$$g_{lm} \sum_n a_n I^n_{lm'} = c^l_{m'm} , \tag{5.16}$$

where $I^n_{lm'}$ are computed for each S_n as in (5.7). Thus, with a sufficient number of equations, the factors $g_{lm} a_n$ can be obtained. From these the ratios of the coefficients a_n – the values of these coefficients up to a common constant – can be obtained. Of course, this common constant is unimportant, since multiplying the Gaussian surface density or the scattering law by a constant corresponds to scaling the size or the brightness of the surface uniformly. It should be noted, however, that if the coefficients a_n depend on the phase angle α, the common constant will become an unknown function of α, so the equations will have to be solved using a fixed value of α.

Since the coefficients a_n are now known, the coefficients g_{lm} can be computed as well. A sufficient number of equations is obtained when l is large enough. This naturally means that there have to be nonzero coefficients $c^l_{m'm}$ at the chosen value l; i.e., the surface has to be sufficiently complicated if the coefficients a_n depend on the phase angle. If a_n are taken to be constants, observations at several phase angles can be used.

Thus, if the functions S_n in the scattering law are known, their relative contributions and the Gaussian surface density can be obtained from the observational data. In practice, the Gaussian surface density often appears in realistic scattering laws in the form of a product with the albedo ϖ_0, since the effects of these two functions on the observed brightness are of a similar nature. In this case, we obtain the product $G\varpi_0$ in the inversion, and the separation of the two functions is impossible. In the next section we shall show that if the functional form of the scattering law is in a suitable way more complicated than the aforementioned linearly albedo-dependent form, the Gaussian surface density can be uniquely determined. The possibility of such discrimination is because the surface of an asteroid is almost certainly covered by small regolith particles whose collective scattering properties differ from those of a consolidated, smooth surface. Therefore, the scattering law need not always be linear in albedo, as in the case of a "painted" surface in which "paint" markings determine albedo and the dependence is thus linear, as in Russell's paper. The greater the value of the single scattering albedo, the greater the effects of multiple scattering between the surface particles, and thus the larger the deviation from the linear albedo dependence.

5.2. *Discriminating albedo variegation*

Consider now a scattering law of the form

$$S(\mu, \mu_0, \alpha; \varpi_0(\vartheta, \psi)) = f(\alpha; \varpi_0(\vartheta, \psi)) \sum_n a_n S_n(\mu, \mu_0, \alpha) , \tag{5.17}$$

the coefficients a_n and the albedo ϖ_0 being unknown. This law is otherwise the same as (5.15), except that it has been multiplied by a function dependent on the location on the surface. At least one of the coefficients a_n must be independent of the phase angle α. Let a_0 be such a coefficient. Let us write

$$G(\vartheta, \psi) \, f(\alpha_i; \varpi(\vartheta, \psi)) = \sum_{lm} b_{lm} \, Y^m_l(\vartheta, \psi) \tag{5.18}$$

at a fixed phase angle α_i. The equations of the previous section apply also here when b_{lm} replaces g_{lm}, so the ratios a_n/a_0 and thus the function $a_0 G(\vartheta,\psi) f(\alpha_i;\varpi_0(\vartheta,\psi))$ are obtained. In practice, at least two coefficients a_n have to be independent of the phase angle, so from the ratios of all the coefficients a_n obtained at several phase angles the α-independent coefficients can be deduced.

If the aforementioned function has been obtained for several values α_n of the phase angle, we find functions

$$F_{nk}(\vartheta,\psi)=\frac{f(\alpha_n;\varpi_0(\vartheta,\psi))}{f(\alpha_k;\varpi_0(\vartheta,\psi))}. \tag{5.19}$$

If the functional form $f(\alpha;\varpi_0)$ is now known, the functions $\varpi_0(\vartheta,\psi)$ can be determined from (5.19) if the form of the function f is suitable (necessarily not merely linear in albedo). Thus we need two phase angles to solve for the albedo distribution in the ideal case.

Consider the fictitious example

$$f(\alpha;\varpi_0)=\varpi_0(1+\varpi_0\cos\alpha). \tag{5.20}$$

The albedo distribution ϖ_0 can be determined using (5.19) with the result

$$\varpi_0(\vartheta,\psi)=\frac{F_{nk}(\vartheta,\psi)-1}{\cos\alpha_n-F_{nk}(\vartheta,\psi)\cos\alpha_k}. \tag{5.21}$$

After ϖ_0 has been computed, the Gaussian surface density can be obtained by substituting ϖ_0 back into the function f.

The other "albedo-separable" form of the scattering law is

$$S(\mu,\mu_0,\alpha;\varpi_0(\vartheta,\psi))=\sum_n P_n(\varpi_0(\vartheta,\psi))\,S_n(\mu,\mu_0,\alpha). \tag{5.22}$$

The Lumme-Bowell law, for example, can be written in this form, and the resulting functions $P_n(\varpi_0)$ are suitable for separation. Let us now write

$$G(\vartheta,\psi)\,P_n(\varpi_0(\vartheta,\psi))=\sum_{lm} b_{lm}^n\,Y_l^m(\vartheta,\psi). \tag{5.23}$$

Now we obtain, as with Eq. (5.16),

$$\sum_n b_{lm}^n\,I_{lm'}^n=c_{m'm}^l. \tag{5.24}$$

As before, the coefficients b_{lm}^n are the solutions of sets of simultaneous linear equations. Now, however, there is a maximum of $2l+1$ equations that can be used to find the coefficients b_{lm}^n using one phase angle. With the lowest values of l this is not enough to define the coefficients uniquely. Therefore the observations made at a single phase angle can be explained by infinitely many different combinations of the functions GP_n. Thus more than one phase angle must be used in the inversion, and the functions P_n have to be independent of the phase angle. If they are of the form $P_n(\alpha,\vartheta,\psi)$, a unique solution is impossible.

Using all the m'-values and a sufficient number of phase angles the functions $G(\vartheta,\psi)\,P_n(\varpi_0(\vartheta,\psi))$ can thus be obtained. As with (5.16), the ratios of the functions P_n can now be computed, and one function dependent on the location will be left unknown. If we denote

$$\tilde{P}_n(\vartheta,\psi)\equiv\frac{P_n(\varpi_0(\vartheta,\psi))}{P_1(\varpi_0(\vartheta,\psi))}, \tag{5.25}$$

we shall be able to determine the functions $\tilde{P}_n$ and the product GP_1.

To determine the Gaussian surface density G, the forms of the function P_1 and another function P_n as functions of albedo ϖ_0 must be known, and they have to be suitable, as were the forms $f(\alpha;\varpi_0)$ in connection with (5.19). In the same manner as in the case of (5.19), the albedo distribution ϖ_0 can be found using the ratio (5.25), and thus the functions P_1 and G can be determined. Other functions P_k max then be of the "direct" form $P_k(\vartheta,\psi)$, but, as already shown, at least some part of the location dependence in the scattering law must be introduced by means of physical parameters.

The separation methods presented here may also be implemented in the inversion methods using sets of simultaneous equations; for the determination of all the functions involved, there must be a sufficient number of equations (i.e., observations at different geometries).

The choice of the functions S_n and the functional forms $P_n(\varpi_0)$ is best made by using the existing physical scattering models and physically probable functional forms. In this, a sort of "iterative guessing" is likely to be used: the set of functions providing the most consistent solutions is chosen. The simplest model might be a combination of the Lommel-Seeliger and Lambert laws, whose relative contributions and the product of the albedo variegation and the Gaussian surface density would be obtained in the inversion. With a large number of observed light curves, the separation of the albedo variegation could be pursued by using more complicated scattering laws.

6. Shape construction from the Gaussian surface density

As can be seen from Eq. (2.4) and (2.6) or from Appendix C, the analytical expressions between the radius vector and the Gaussian surface density of a surface lead to nonlinear partial differential equations that cannot generally be solved. Also, it has been known for some time that in a strictly convex case the Gaussian surface density uniquely determines the surface (Minkowski 1903; Nirenberg 1953), but unfortunately the proof is not directly suitable for constructing the shape when G is known. Thus, indirect numerical methods must be used instead. Hitherto, a proper numerical solution of the problem has been unknown.

We have developed two methods for approximating the solution. Both use Minkowski's proof of the uniqueness between a strictly convex surface and its Gaussian surface density. Minkowski (1903) showed that the so-called mixed volume, constructed from G and the support function, a function uniquely describing a surface, reaches its minimum when the support function and the Gaussian surface density correspond to the same surface. Thus the problem can be represented as a constrained minimization problem to be iteratively solved. Constrained minimization is a problem often encountered in various situations; our methods for implementation are partly based on common procedures (see, e.g., Gill et al. 1981), although they are specifically tailored to the problem at hand.

In what follows, we describe the theoretical principles of the two surface reconstruction methods. The actual algorithms and detailed discussion will be presented elsewhere. First, we describe the concepts of support function and mixed volume, central to Minkowski's minimization theorem.

Let $\boldsymbol{n}(\vartheta,\psi)=(\sin\vartheta\cos\psi,\ \sin\vartheta\sin\psi,\ \cos\vartheta)^{\mathrm{T}}$ be the unit outward normal and $\boldsymbol{r}(\vartheta,\psi)$ the radius vector of the surface. The inner product $\boldsymbol{n}\cdot\boldsymbol{r}$ is called the support function ϱ of the surface:

$$\varrho(\vartheta,\psi)=\boldsymbol{n}(\vartheta,\psi)\cdot\boldsymbol{r}(\vartheta,\psi). \tag{6.1}$$

Thus $\varrho(\vartheta,\psi)$ is the distance, measured from the origin, of the tangent plane on the surface at point (ϑ,ψ). The radius vector of a

surface is uniquely obtainable from the support function as shown in Appendix C.

The mixed volume $\tilde{V}(R, S)$ of two strictly convex bodies R and S is defined here as

$$\tilde{V}(R,S)=\frac{1}{3}\int_0^{2\pi}\int_0^{\pi}\varrho_R(\vartheta,\psi)\,G_S(\vartheta,\psi)\,\sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\psi\,, \tag{6.2}$$

where the indices of the functions refer to the corresponding bodies. Originally, Minkowski's concept of mixed volume involved three bodies (one support function and two Gaussian surface densities), but for illustrating the surface reconstruction method, only two bodies are needed. Using Eq. (6.7), introduced later in this section, it can directly be seen that the mixed volume is translation invariant; i.e., the origin for the support function can be arbitrarily chosen. It is also straightforward to see that the volume of a body S is $\tilde{V}(S, S)$, and, if we denote the unit sphere by U, the surface area of S is $3\,\tilde{V}(U, S)$.

Now let the volume of R in (6.2) be 1 (or some other positive constant). Minkowski (1903) has shown, that $\tilde{V}(R, S)$ reaches its minimum exactly when R is homothetic with S; i.e., $\varrho_R(\vartheta, \psi)$ describes (up to a scaling factor and translation) the same body as $G_S(\vartheta, \psi)$. Thus, by solving a constrained minimization problem to recover the support function, the surface corresponding to a known G can be reconstructed.

Mathematically, both of our surface reconstruction methods can be represented in a discrete form as procedures minimizing the inner product $\langle \boldsymbol{x}, \boldsymbol{g}\rangle$ of vectors in R^n-space, $\boldsymbol{g}$ being known from the Gaussian surface density and $\boldsymbol{x}$ being the unknown vector representing the support function. The discretization of a continuous function into a vector is done either by using the coefficients of a truncated Laplace series or by using a polyhedron approximating a continuous surface. The minimization procedure is subject to the condition $V(\boldsymbol{x}) = 1$, V being the volume computed from $\boldsymbol{x}$. In practice, the equivalent procedure of maximizing $V(\boldsymbol{x})$ while staying on the hyperplane $\langle \boldsymbol{x}, \boldsymbol{g}\rangle$ = constant is computationally more efficient.

In the first of our methods, which we call the method of the generating function, the support function is expressed as a truncated Laplace series, whose coefficients we then solve. The truncation point is the same as that of the Laplace series describing the Gaussian surface density. The coefficients of degree 1 can be ignored, since they represent only a translation. In this case (6.2) is written as

$$\tilde{V}(R,S)=\sum_{lm}\varrho_{lm}^{(R)}\,a_{lm}^{(S)}\,, \tag{6.3}$$

where

$$a_{lm}^{(S)}=\frac{1}{3}\int_0^{2\pi}\int_0^{\pi}G_S(\vartheta,\psi)\;Y_l^m(\vartheta,\psi)\,\sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\psi\,, \tag{6.4}$$

and the coefficients $\varrho_{lm}^{(R)}$ are numerically solved for by minimizing (6.3) iteratively. We cannot state a strict proof of convergence of iteration, but some mathematical considerations give a good reason to expect that the method usually yields a good approximation. The simulations we have performed also support this expectation. The method of the generating function is in practice very useful, both because it is very fast to implement and because the result is given in a practical form as coefficients of a Laplace series.

In the second method, which we call the polyhedron method, the Gaussian surface density is discretized so that the areas and outward normals of the facets of a polyhedron approximating the shape are known. Previously, there have been some studies along the same lines concerning the problem of reconstructing a polyhedron from its extended Gaussian image (Little 1983), although these studies have not been formulated rigorously. The mixed volume in the discrete case, corresponding to Eq. (6.2), is written as

$$\tilde{V}(R,S)=\frac{1}{3}\sum_{j=1}^{n}l_j^{(R)}\,A_j^{(S)}\,, \tag{6.5}$$

where A_j is the area of a facet and l_j is its rectangular distance from the origin. We now search for the set of distances $l_j^{(R)}$ of the facets, measured from the origin, minimizing (6.5) with the same constraints as mentioned in connection with (6.2).

The discretized G naturally has to describe a convex polyhedron. This means that if the unit outward normal of a facet multiplied by the area of the facet is denoted by the vector $\boldsymbol{a}_j$, the equation

$$\sum_{j=1}^{n}\boldsymbol{a}_j=\boldsymbol{0} \tag{6.6}$$

must hold true. In the continuous case,

$$\iint_{S^2}G(\boldsymbol{n})\,\boldsymbol{n}\,\mathrm{d}\sigma=\boldsymbol{0}\,, \tag{6.7}$$

where $\boldsymbol{n}$ is the unit outward normal, S^2 denotes the unit sphere of the Gaussian image, and $\mathrm{d}\sigma$ its surface differential. This results from the geometrical requirement, mentioned in the end of Sect. 4, that the projected areas of a surface, viewed from opposite directions, be equal.

A discretization fulfilling condition (6.6) is

$$\boldsymbol{a}_j=\iint_{U_j}G(\boldsymbol{n})\,\boldsymbol{n}\,\mathrm{d}\sigma \tag{6.8}$$

integrated over a portion U_j of the unit sphere. When the U_j are suitably chosen, the polyhedron approximates the original surface. Our iteration procedure demonstrably converges towards the correct solution, so the larger the number of facets, the better the approximation. Also, according to the Brunn-Minkowski theorem (see, e.g., Grünbaum 1967), the set of the distances $\{l_j^{(R)} \mid \tilde{V}(R, R) \geqq 1\}$ in R^n is convex, so a local minimum of (6.5) is the global minimum, which is advantageous in solving the minimization problem: the minimum to be determined is the only minimum.

It is an important fact that the radius vector $\boldsymbol{r}$ of the surface depends continuously on the Gaussian surface density, so small errors in G, in the sense of L^2 norm (square integration norm), remain small in the evaluated $\boldsymbol{r}$. The numerical examples we have computed show this clearly. Thus the problem of inferring the shape from the Gaussian surface density does not produce stability problems, in contrast to the problem of deducing the Gaussian surface density from the lightcurves. In fact, errors that are no longer very small in G may still correspond to practically negligible errors in the evaluated shape. Because of this property, shape determination is in a way more stable than the determination of the albedo distribution. Thus it is "safer" to assume that light-variations are caused mainly by the shape rather than by the albedo distribution, if such an assumption has to be made. However, as explained in Sect. 4, coefficients of degree 1 in a Laplace series obtained by inversion can be explained only by albedo variation or non-convex shape.

7. Numerical simulations

In this section we briefly describe some simulations with synthetic observational data. In the next paper we shall study the properties related to practical observations more closely. Our test object here is a strictly convex body without albedo variegation, shown as viewed from three mutually perpendicular directions in Fig. 1. The direction of the rotation axis is pointing upwards in the plane of the paper in Figs. 1a and b; in Fig. 1c the body is viewed from directly above so that the viewing directions corresponding to Figs. 1a and b are, respectively, directly below and on the right-hand side of Fig. 1c.

In Figs. 1a and b the solar phase angle is 60°, the illumination direction being perpendicular to the rotation axis; in Fig. 1c the illumination direction is the same as in Fig. 1a. For purposes of illustration, the scattering law in the images is the Lommel-Seeliger law. The boundary curve representing the shadowed part of the limb, which would be seen if the phase angle were zero, is separately drawn. The "contours" on the surfaces appear because of the resolution of the images: the surfaces were drawn by computing points, typically a few degrees apart from each other, on a grid on the Gaussian unit sphere. Thus, along a "contour", the latitude or longitude of the direction of the surface normal is constant. Because of the viewing/illumination geometries, in Figs. 1a and b the longitudinal lines are seen more clearly than the latitudinal lines, whereas in Fig. 1c the latitudinal lines appear more distinctly. If, e.g., the phase angle in Fig. 1a or b were 12° larger, the terminator would be the longitudinal contour next to the terminator illustrated.

Lightcurves produced by the model in various observing geometries can easily be computed using Eqs. (5.6) or (5.8) (the Gaussian surface density of the object is explicitly known). The scattering law chosen was a combination of Lambert's law and the Lommel-Seliger law such that the ratio of the former to the latter (we call this the L/LS ratio) was 0.2. A moderate amount of noise (about two percent on average) was added to the lightcurves. The exact Laplace series of the Gaussian surface density of this object contains terms to the 10th degree.

Two inversion techniques were used. One method was based on the orthogonality property of the elements of the rotation matrices of spherical harmonics, as explained in Sect. 5: Eqs. (5.13), (5.14), etc. The fixed phase angle was chosen to be 14°. The first few coefficients of the three-dimensional Fourier series describing $L(\kappa, \varepsilon, \delta)$ were approximated by using 16 lightcurves obtained at different well distributed observation geometries (equally spaced intervals covering the κ- and ε-ranges). The Laplace series describing the Gaussian surface density solution was then truncated at order 3 and degree 3 because the Fourier series naturally gave no information on terms of higher degree. The L/LS ratio of the scattering law used in inversion was 0.4, thus being slightly different from the ratio used in creating the lightcurves. The method used in obtaining the shape from the Gaussian surface density was the method of the generating function (see Sect. 6). The shape solution is shown in Fig. 2, the viewing/illumination directions being the same as in Fig. 1.

The other method was based on sets of Eqs. (5.11) that can be used when the Fourier coefficients of each lightcurve are known. [One could, of course, use the single brightness values as described in the beginning of Sect. 5.1. However, the form (5.11) is much more compact, since it divides a large set of equations into smaller sets.] In this case a data set of "real" observing geometries was used. This set consisted of 12 lightcurves whose observing geometries were the same as those of 39 Laetitia described in

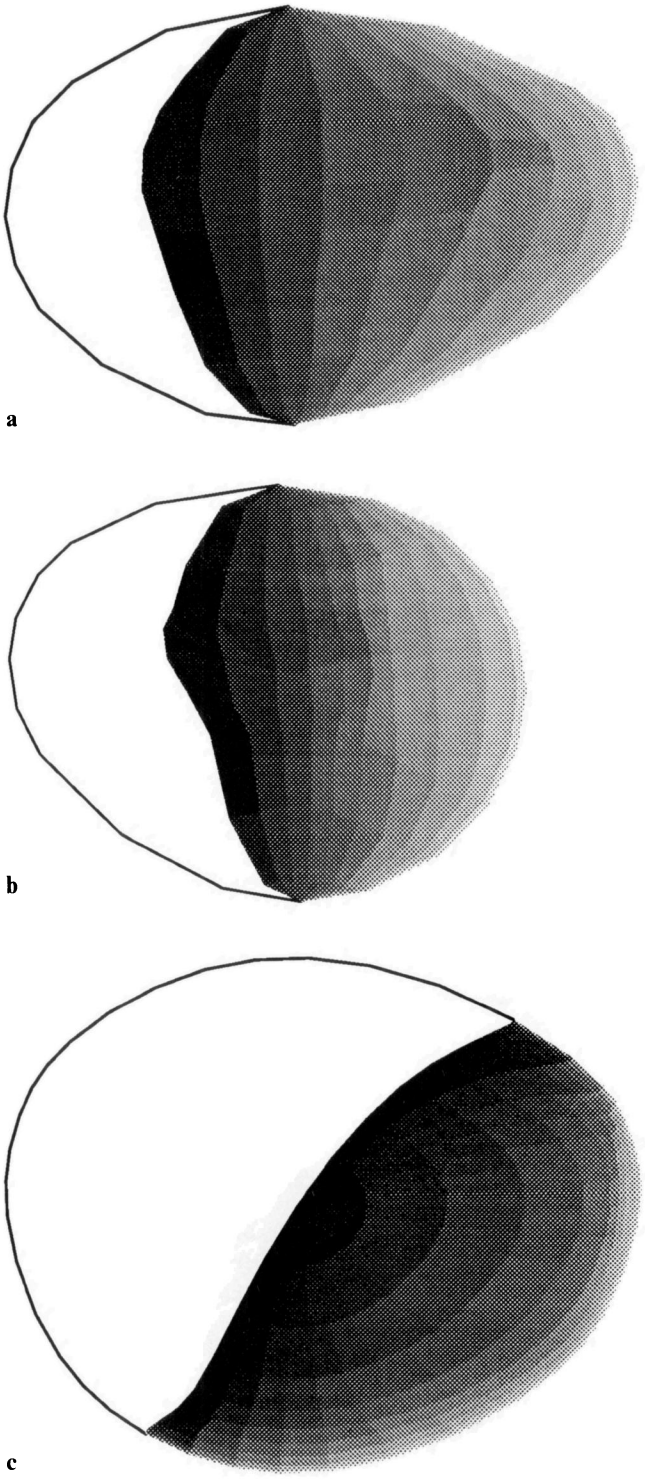


Fig. 1a–c. The model used to produce synthetic lightcurves, shown as viewed from three mutually perpendicular directions. The shadowed part of the limb is also shown. The direction of the rotation axis is pointing upwards in the plane of the paper in **a** and **b**; in **c** the body is viewed from directly above such that the viewing directions corresponding to **a** and **b** are, respectively, directly below and on the right-hand side of **c**. In **a** and **b** the solar phase angle is 60°, the illumination direction being perpendicular to the rotation axis; in **c** the illumination direction is the same as in **a**

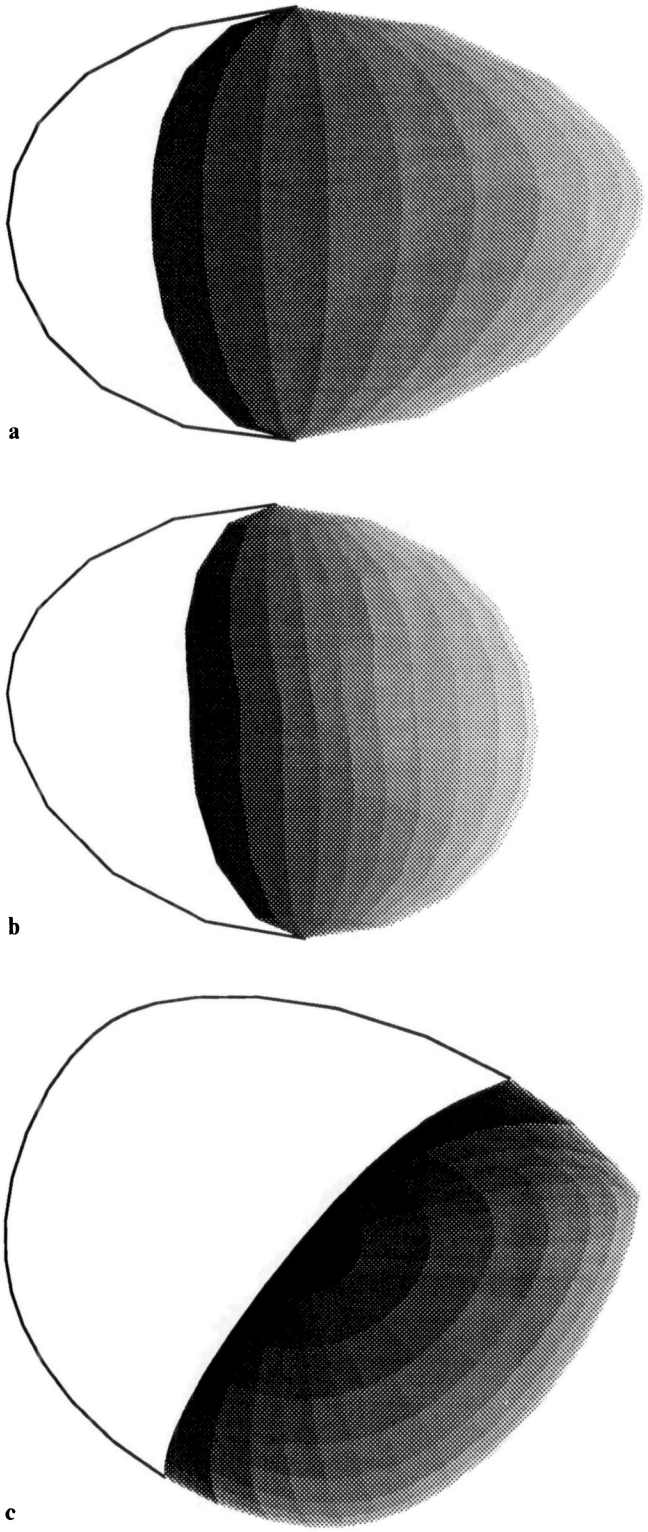


Fig. 2a–c. The shape solution obtained from 16 lightcurves at well distributed observing geometries at solar phase angle 14° and using the "orthogonality method". The viewing/illumination directions are as in Fig. 1

Lumme et al. (1990); the direction of the rotation axis was chosen to be the pole solution presented in that paper. With no a priori knowledge, the solutions for the higher-degree coefficients of sets of equations of this kind tend, in particular, to be unstable, as is common in ill-posed problems. The stabilizing scheme was chosen to be the method of statistical inversion. The a priori assumption used imposed no great limitations; it simply restricted the ranges of the unknown coefficients to correspond to a rapidly converging Laplace series. The Laplace series of the solution was truncated at degree 4 and order 4, and the L/LS ratio used in inversion was 0.3. The shape solution is shown in Fig. 3, the viewing/illumination directions corresponding to those in Figs. 1 and 2.

Figures 2 and 3 rather well reproduce the coarse-scale global shape of the initial object of Fig. 1. The surface of the object in Fig. 2 appears to be the smoothest of the surfaces in the figures. This is because its corresponding Laplace series of G obtained in inversion is everywhere positive (as a proper G should be) while it contains terms only to the third degree: a short series of this kind describes a very smooth surface. The surface in Fig. 3, in contrast, contains some local less smooth features such as the "ridge" or "crease" seen in Figs. 3b and c. It is important to note that this is also a result of the truncation of the Laplace series describing the Gaussian surface density. In this case, the solution for G has negative values at some points, which, of course, is erroneous. However, reconstruction of the shape is not precluded because the error is sufficiently small. Thus Fig. 3 contains some local "details" that are merely side effects of the inversion method; the actual result of inversion is the global structure on a rather large scale, because the number of lightcurves used in inversion is rather small.

To quantify the goodness of fit we use two estimators to compare the initial surface with the reproduced surface. The first estimator $\Delta\varrho$ gives the average relative error of the support function when the volume of the initial body is normalized to that of the reproduced body and the centers of mass of the two bodies are translated to coincide. Thus

$$\Delta\varrho = \frac{1}{4\pi} \iint_{S^2} \frac{|\varrho_{\text{in}} - \varrho_{\text{app}}|}{\varrho_{\text{in}}} \, \mathrm{d}\sigma, \tag{7.1}$$

where ϱ_{in} and ϱ_{app} are, respectively, the inital and the approximate support functions and $\mathrm{d}\sigma$ is the surface differential on the Gaussian unit sphere S^2.

The second estimator $\Delta_2\varrho$ is the standard deviation of the first and is computed from

$$\Delta_2\varrho = \left[\frac{1}{4\pi} \iint_{S^2} \left(\frac{|\varrho_{\text{in}} - \varrho_{\text{app}}|}{\varrho_{\text{in}}} - \Delta\varrho \right)^2 \mathrm{d}\sigma \right]^{1/2}. \tag{7.2}$$

The two estimators are not symmetric with respect to the support functions used. Therefore, with $\Delta\varrho$ and $\Delta_2\varrho$ the initial surface relative to which the estimators were computed must also be defined. In what follows, we use the phrase "B compared to A" to denote that the support functions ϱ_{in} and ϱ_{app} are related to A and B, respectively. In practice, the difference resulting from the choice of the support function, relative to which the values are computed, is usually negligible.

Comparing the computed shape in Fig. 2 to that of the model in Fig. 1 we have $\Delta\varrho = 0.016$, $\Delta_2\varrho = 0.009$; for Fig. 3 the corresponding values are $\Delta\varrho = 0.023$, $\Delta_2\varrho = 0.013$. To establish a scale for these values, we note that when the triaxial ellipsoid best describing the lightcurves produced by the initial shape in Fig. 1 (the axes of the ellipsoid are 0.7, 1.0, and 1.2) is compared to this shape in the same manner as above, we obtain $\Delta\varrho = 0.040$ and

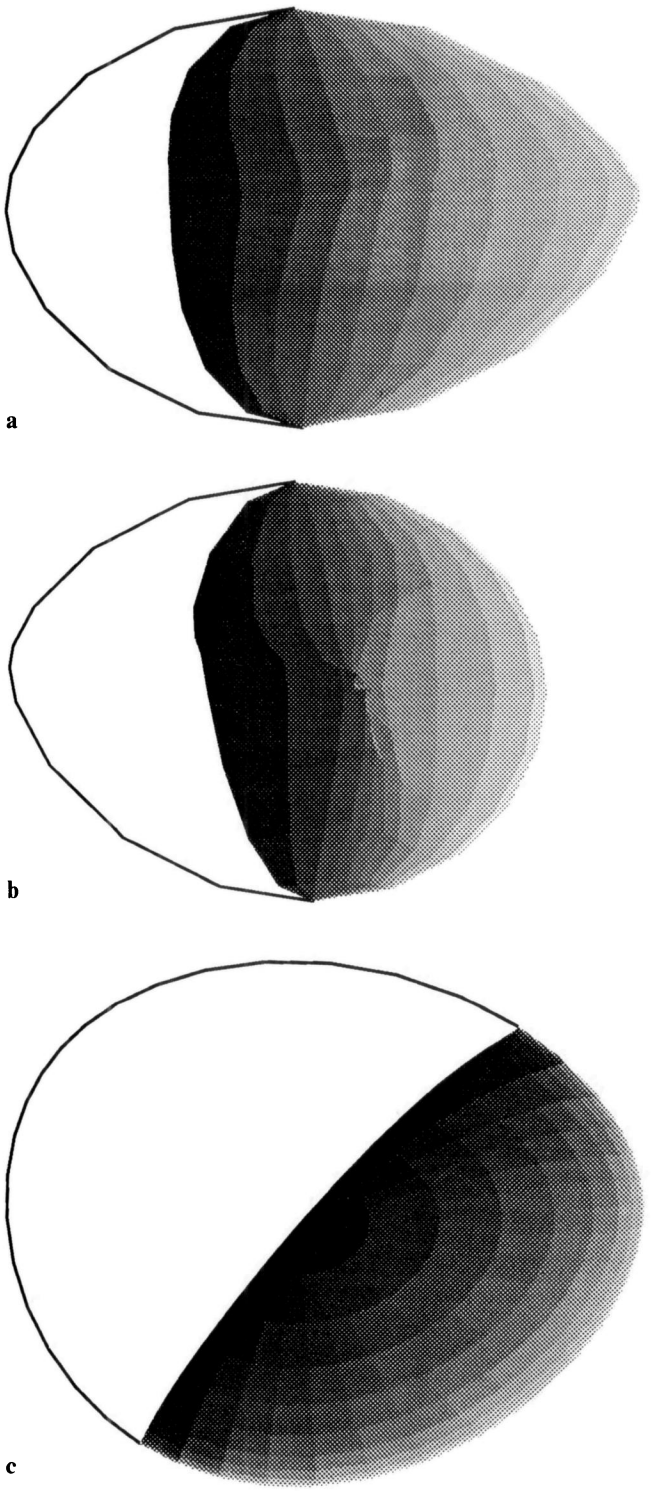


Fig. 3a–c. The shape solution obtained from 12 lightcurves at the input observing geometries and using sets of equations solved by statistical inversion. The viewing/illumination directions correspond to those in Figs. 1 and 2

$\Delta_2\varrho = 0.026$. If a sphere is compared to this ellipsoid, the values $\Delta\varrho = 0.12$ and $\Delta_2\varrho = 0.08$ are obtained. If a sphere is compared to a more elongated ellipsoid (axes 0.5, 1.0, and 1.5), we have $\Delta\varrho = 0.22$ and $\Delta_2\varrho = 0.15$. Thus the results of inversion can also quantitatively be seen to be somewhat better than the best ellipsoid fit.

8. Conclusions and discussion

We have studied the photomorphography problem in detail, and we have derived theoretical inversion methods that can produce more information from lightcurves than Russell's (1906) results might lead one to believe. The key factor is naturally the number of lightcurves obtained at different observing geometries. In practice, the method of rotation matrices described in Sect. 5 requires more lightcurves for consistent implementation than the methods using sets of equations, such as (5.11), combined with statistical inversion. With these methods it is always possible to obtain a solution even with only a few lightcurves, but if the amount of observational data is not sufficient, the solution obtained may be very uncertain especially for the higher-order coefficients. However, numerical simulations show that Laplace series describing "well-behaved" surfaces often converge rapidly, allowing truncation of the series at an early stage.

From various simulations that we have performed it would seem that using at least, say, ten well-distributed lightcurves, a reasonable large-scale estimation of the shape is often possible. We estimate that for a shape solution that can more reliably be expected to be near the correct one, there should preferably be at least about twenty different lightcurves. It should also be noted that because lightcurves measured at nonzero solar phase angles are necessary for our analysis, there should be lightcurve observations as far away from opposition geometry as possible. In the next paper, we shall study the effect of solar phase angle on the obtainable information more closely.

To obtain information on a scattering law of the form (5.17) or (5.22) and to properly separate the effects of albedo variegation from those of shape, the number of lightcurves must be many times larger. Therefore, at least with the existing data, one must decide whether the function obtained as a solution describes albedo or shape features. There are some indirect indicators: as mentioned in Sect. 4, certain coefficients obtained in inversion indicate albedo variegation. Also infrared lightcurves can indicate albedo variegation when compared to lightcurves in visible light, because a thermal-radiation lightcurve behaves differently from that measured in visible light in the face of albedo variegation (Magnusson et al. 1989).

Sufficiently accurate determination of the observing geometry is also an important factor. The pole position of the asteroid can often be determined to an accuracy of 20° or so. In our next paper concerning photomorphography we shall show that inaccuracies within this limit probably do not produce significant errors at least in the coefficients of the first few degrees of the Laplace series describing the Gaussian surface density. Thus the shape solution is rather insensitive to small errors in the pole position. We shall also show – and this is intuitively quite clear – that the sidereal rotation period of the asteroid must be very precisely known, if it is to be used in computing the absolute rotational phases. Typically, the period must be known to an accuracy of one tenth of a second to make a useful determination of the rotational phases of the observations in lightcurves obtained over many years. Such an accuracy is probably seldom obtainable a priori. One possibility is

to use photometric features or other properties of lightcurves to adjust the location of each lightcurve in rotational phase. Another method, as mentioned in Sect. 2, is to make use of a series of small deviations from a previously determined period to choose the period that gives the best fit in the photomorphographic analysis.

At present, there are few asteroids that have been well enough observed. Suitable candidates for photomorphographic analysis are, for example, 39 Laetitia and 16 Psyche. There are also some asteroids for which sufficient lightcurves can be attained if their next few apparitions are properly observed; it might be worthwhile to look through asteroid lightcurve data collected so far and search for useful candidates of which additional observations could be made. Even though lightcurve data for a given asteroid are insufficient for shape determination, they may perhaps be used in conjunction with the inversion methods we have shown to obtain a more accurate value of the asteroid's rotation period or pole position.

Acknowledgements. We would like to thank Berthold K.P. Horn, William Irvine, Hannu Karttunen, and Karri Muinonen for useful comments and suggestions. EB was supported largely by NASA grant NAGW-1470.

Appendix A: coordinate systems

We describe here the four coordinate systems used in the text. The systems $C2$ and $C3$ are not coordinate systems in the usual sense of describing points in space; they are used to describe entire observing geometries with four variables.

System C1

We use the following notations in the $C1$ Cartesian coordinate system fixed to the asteroid, the z-axis being the rotation axis (Fig. A1).

$\boldsymbol{E}$ = a unit vector toward the observer (Earth),
$\boldsymbol{E}_0$ = a unit vector toward the Sun,
$\boldsymbol{n}$ = surface outward normal, a unit vector,
$\mu = \boldsymbol{E} \cdot \boldsymbol{n}$,
$\mu_0 = \boldsymbol{E}_0 \cdot \boldsymbol{n}$,
$\cos \alpha = \boldsymbol{E} \cdot \boldsymbol{E}_0$.

The spherical polar coordinate pairs representing the directions of the Sun and the Earth are (θ_0, φ_0) and (θ, φ). The ranges of the angles are

$$\begin{cases} 0 \leqq \theta,\ \theta_0 \leqq \pi \\ 0 \leqq \varphi,\ \varphi_0 < 2\pi\,. \\ 0 \leqq \alpha \leqq \pi \end{cases}$$

System C2

This system consists of two coupled spherical coordinate systems, the first of which defines the direction of the Earth, as in $C1$. The polar axis of the second system lies in the direction of the Earth, and the direction of the Sun is defined by the spherical polar coordinates (α, γ) of this system (Fig. A2), α being the polar angle. The longitudinal angle γ is the obliquity, its zero longitude being the semicircle formed by the great circle going through the direction of the Earth and through the direction of $(\frac{\pi}{2}, \varphi + \frac{\pi}{2})$ in $C1$. In other words, the obliquity is, with opposite sign, the complement of the angle between the half-plane defined by the line of sight and the spin vector and that defined by the line of sight and the Sun. Using general relations of spherical trigonometry, we have

$$\cos\gamma = \frac{\sin\theta_0 \sin(\varphi_0 - \varphi)}{\sin\alpha},$$

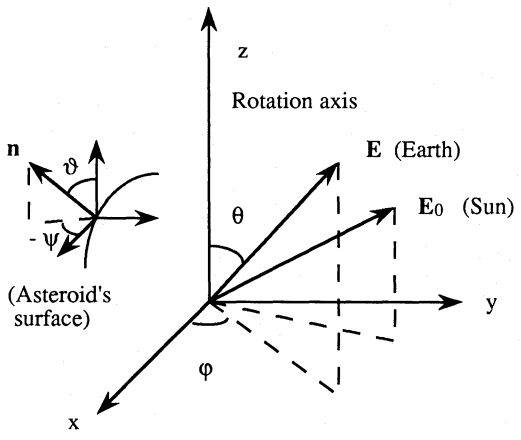
$$\cos\alpha = \cos\theta\cos\theta_0 + \sin\theta\sin\theta_0\cos(\varphi_0 - \varphi)\,. \qquad \text{(A1)}$$

The ranges of the variables are as above, with $-\pi < \gamma \leqq \pi$. The quadrant of γ is chosen using the sign of $\sin\gamma$, given by the sign of $\cos\theta_0 - \cos\alpha\cos\theta$ (this is equal to $\sin\gamma\sin\alpha\sin\theta$).

System C3

This system is derived from the system $C1$ by rotation. The angles δ, ε and κ are the Euler angles (Appendix B) used in transforming the coordinate system $C1$ into a system in which the y-axis lies in the direction of the Earth and the direction of the Sun lies in the xy-plane at an angle α to the direction of the Earth, measured in the positive rotation direction.

Although this system is rather abstract, the variables $\alpha, \delta, \varepsilon$ and κ uniquely define the directions of the Sun and the Earth. Using spherical trigonometry again, we obtain



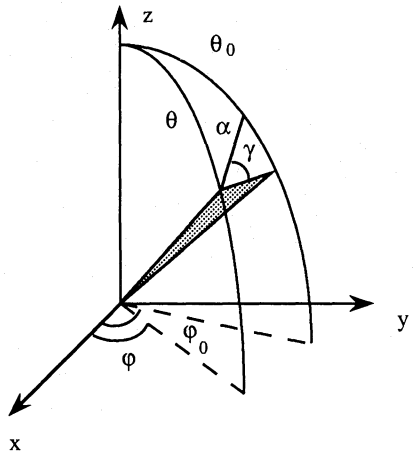


Fig. A1. (*left*) The Cartesian coordinate system $C1$ fixed to the asteroid. A strictly convex surface can be globally parametrized using the spherical polar coordinates (ϑ, ψ) of the direction of the surface normal $\boldsymbol{n}$

Fig. A2. (*right*) The coordinate system $C2$ defining the direction of the Sun by the obliquity γ and the solar phase angle α

$$\delta = \varphi - \frac{\pi}{2} + (2S(\gamma) - 1) \arctan(\cos\theta \tan\beta) + S(\gamma)\pi$$

$$\varepsilon = \arccos(\sin\theta \sin\beta)$$

$$\kappa = S(\gamma)\,\pi - (2S(\gamma) - 1) \arctan\left(\frac{\cot\theta}{\cos\beta}\right), \tag{A2}$$

where $S(\gamma)$ is the unit step function "backwards":

$$S(\gamma) = \begin{cases} 0, \gamma > 0 \\ 1, \gamma \leqq 0 \end{cases}, \tag{A3}$$

$\sin\beta = \cos\gamma$; $\cos\beta \geqq 0$,

the values of the arctangent are in the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$, and the ranges of the coordinate angles are

$$\begin{cases} 0 \leqq \delta, \kappa < 2\pi, \\ 0 \leqq \varepsilon \leqq \pi. \end{cases}$$

In the special case $\theta = \theta_0 = \frac{\pi}{2}$ we have $\delta = \varphi - \frac{\pi}{2}$ and $\kappa = 0$; at opposition ($\alpha = 0$) the values $\delta = \varphi$, $\varepsilon = \theta - \frac{\pi}{2}$, $\kappa = -\frac{\pi}{2}$ can be chosen. Also, to illustrate the geometry of δ, ε and κ, it is interesting to note that if the rotation axis and the orbital plane of the asteroid lie in the plane of the ecliptic; i.e., $\varphi - \varphi_0$ equals 0 or π ($\gamma = \pm\frac{\pi}{2}$), the angle ε will be $\frac{\pi}{2}$ and the angles δ and κ will be, respectively, φ and θ with phase shifts of magnitude $\frac{\pi}{2}$.

Surface coordinate System C4

This system parametrizes a point on a surface using the direction of the surface outward normal at that point, given by the spherical polar coordinates (ϑ, ψ) (Fig. A1). For a strictly convex surface, this mapping (the Gaussian image) from the surface onto a unit sphere provides a global parametrization.

Appendix B: spherical harmonics, Euler angles and rotation matrices

The spherical harmonics $Y_l^m(\vartheta, \psi)$ are here defined as

$$Y_l^m(\vartheta, \psi) = P_l^m(\cos\vartheta)\, e^{im\psi}, \tag{B1}$$

where P_l^m is an associated Legendre polynomial. Normalization coefficients or "phase factors" such as $(-1)^m$ are here not included in the functions themselves. For practical purposes, a useful definition is

$$P_l^m(\cos\vartheta) = \sum_{n=0}^{l-m} A_{lm}^n \cos^n\vartheta \sin^m\vartheta, \tag{B2}$$

where

$$A_{lm}^n = \frac{1}{l-m}\left[(2l-1)\, A_{l-1\,m}^{n-1} - (l+m-1)\, A_{l-2\,m}^n\right] \tag{B3}$$

for $l > m$ and $l > 1$, and

$$A_{ll}^0 = (2l-1)\, A_{l-1\,l-1}^0; \qquad A_{00}^0 = 1. \tag{B4}$$

Only such A_{lm}^n for which n is of the same parity as $l-m$ are nonzero. Also, note carefully that for negative m-indices

$$P_l^{-m} = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m. \tag{B5}$$

Throughout our presentation we use the imaginary exponent notation in conjunction with spherical harmonics because of its convenience. For a real-valued Laplace series, Eq. (B5) relates the coefficients with negative m to those with positive m; thus only the $m \geqq 0$-coefficients are of interest. The formalism can, of course, quite readily be transformed into the manifestly real-valued notation involving sines and cosines, if needed.

Spherical harmonics satisfy the orthogonality relation

$$\int_0^{2\pi}\int_0^{\pi} Y_l^{m*}(\vartheta, \psi)\; Y_j^n(\vartheta, \psi) \sin\vartheta \,d\vartheta\, d\psi = N_{lm}\delta_{jl}\delta_{mn}, \tag{B6}$$

where

$$N_{lm} = \frac{4\pi}{2l+1}\frac{(l+m)!}{(l-m)!}, \tag{B7}$$

so that if a function $f(\vartheta, \psi)$ is expanded as a Laplace series

$$f(\vartheta, \psi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} f_{lm}\, Y_l^m(\vartheta, \psi), \tag{B8}$$

the coefficients f_{lm} are

$$f_{lm} = N_{lm}^{-1}\int_0^{2\pi}\int_0^{\pi} Y_l^{m*}(\vartheta, \psi)\, f(\vartheta, \psi) \sin\vartheta \,d\vartheta\, d\psi. \tag{B9}$$

All rotations in three-dimensional space can be parametrized by Euler angles, here defined as follows. The rotation $R(\gamma, \beta, \alpha)$ of a coordinate system involves three successive rotations:

1. A rotation through angle α about the z-axis in the positive rotation direction. As a result of such a rotation, a right-handed screw would travel in the positive direction along the z-axis.
2. A rotation through angle β about the new y-axis in the positive rotation direction.
3. A rotation through angle γ about the newest z-axis in the positive rotation direction.

Satisfying the condition (3.4), spherical harmonics transform under the rotation $R(\gamma, \beta, \alpha)$ as

$$Y_l'^m(\vartheta, \psi) = \sum_{m'=-l}^{l} Y_l^{m'}(\vartheta, \psi)\, D_{m'm}^{(l)}(\gamma, \beta, \alpha), \tag{B10}$$

where $Y_l'^m$ is the form of the function in the new coordinate system, and the element $D_{m'm}^{(l)}$ of a rotation matrix $D^{(l)}$ (cf. Weissbluth 1980, noting the difference in the normalization convention) is

$$D_{m'm}^{(l)}(\gamma, \beta, \alpha) = e^{im'\gamma}\, d_{m'm}^{(l)}(\beta)\, e^{im\alpha}, \tag{B11}$$

where

$$d_{m'm}^{(l)}(\beta) = \sum_{\lambda=0}^{l+m} \cdot \frac{(-1)^{\lambda+m'-m}(l+m)!\,(l-m')!}{\lambda!\,(l+m-\lambda)!\,(l-m'-\lambda)!\,(m'-m+\lambda)!} \cdot \left(\cos\frac{\beta}{2}\right)^{2l+m-m'-2\lambda}\left(\sin\frac{\beta}{2}\right)^{m'-m+2\lambda}. \tag{B12}$$

A term in the sum vanishes if an argument of a factorial in the denominator is negative. A relationship between spherical harmonics and the elements of the rotation matrices is

$$Y_l^m(\vartheta, \psi) = D_{0m}^{(l)}(0, \vartheta, \psi). \tag{B13}$$

The elements of the rotation matrices satisfy an orthogonality relation

$$\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi} D^{(l)*}_{m'm}(\gamma,\beta,\alpha)\, D^{(j)}_{nk}(\gamma,\beta,\alpha)\,\sin\beta\,\mathrm{d}\alpha\,\mathrm{d}\beta\,\mathrm{d}\gamma = N^l_{m'm}\,\delta_{lj}\,\delta_{m'm}\,\delta_{mk}\,, \tag{B14}$$

where

$$N^l_{m'm} = \frac{8\pi^2}{2l+1}\,\frac{(l-m')!\,(l+m)!}{(l+m')!\,(l-m)!}\,. \tag{B15}$$

Appendix C: support function and Gaussian surface density

By differentiating the support function (6.1), an expression for the radius vector of the surface can be directly obtained. The partial derivative of the support function with respect to ϑ is

$$\varrho_\vartheta(\vartheta,\psi) = (\boldsymbol{n}\cdot\boldsymbol{r})_\vartheta = \boldsymbol{n}_\vartheta\cdot\boldsymbol{r}\,, \tag{C1}$$

since $\boldsymbol{n}\cdot\boldsymbol{r}_\vartheta = 0$ (the vectors $\boldsymbol{n}$ and $\boldsymbol{r}_\vartheta$ are, naturally, orthogonal). Correspondingly $\varrho_\psi = \boldsymbol{n}_\psi\cdot\boldsymbol{r}$, so using the support function and its derivatives a set of equations for the components of the radius vector is obtained. Since the vectors $\boldsymbol{n}, \boldsymbol{n}_\vartheta$ and $\boldsymbol{n}_\psi$ are orthogonal, this set of equations can be written, using an orthogonal matrix, as

$$\boldsymbol{\rho} = M\,\boldsymbol{r}\,, \tag{C2}$$

where

$$\boldsymbol{\rho}^{\mathrm{T}} = (\varrho_\vartheta,\ \varrho_\psi/\sin\vartheta,\ \varrho) \tag{C3}$$

and

$$M = \begin{pmatrix} \boldsymbol{n}_\vartheta^{\mathrm{T}} \\ \boldsymbol{n}_\psi^{\mathrm{T}}/\sin\vartheta \\ \boldsymbol{n}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} \cos\vartheta\cos\psi & \cos\vartheta\sin\psi & -\sin\vartheta \\ -\sin\psi & \cos\psi & 0 \\ \sin\vartheta\cos\psi & \sin\vartheta\sin\psi & \cos\vartheta \end{pmatrix} \tag{C4}$$

and $M^{\mathrm{T}}M = I$ when I denotes a unit matrix. Thus

$$\boldsymbol{r}(\vartheta,\psi) = M^{\mathrm{T}}\,\boldsymbol{\rho}(\vartheta,\psi)\,, \tag{C5}$$

so the shape of the surface can readily be reconstructed from the support function.

Also the Gaussian surface density can, of course, be expressed using the support function. A convenient expression can be obtained by starting from the basic definition of the Gaussian surface density: the area of a differential surface patch on an ovaloid body is the area of the corresponding patch on the unit sphere of the Gaussian image multiplied by the Gaussian surface density. Thus, using Eq. (2.4) and denoting $|\boldsymbol{J}| = J$,

$$G(\vartheta,\psi) = \frac{J\,\mathrm{d}\vartheta\,\mathrm{d}\psi}{\left|\dfrac{\partial}{\partial\vartheta}\dfrac{\boldsymbol{J}}{J}\times\dfrac{\partial}{\partial\psi}\dfrac{\boldsymbol{J}}{J}\right|\mathrm{d}\vartheta\,\mathrm{d}\psi}\,, \tag{C6}$$

so that

$$G(\vartheta,\psi) = \frac{J^4}{(\boldsymbol{J}_\vartheta\times\boldsymbol{J}_\psi)\cdot\boldsymbol{J}}\,, \tag{C7}$$

from which, using Eq. (2.6) in the end, it follows that

$$G(\vartheta,\psi) = \frac{(\boldsymbol{n}\cdot\boldsymbol{r}_{\vartheta\vartheta})\,(\boldsymbol{n}\cdot\boldsymbol{r}_{\psi\psi}) - (\boldsymbol{n}\cdot\boldsymbol{r}_{\vartheta\psi})^2}{\sin^2\vartheta}\,. \tag{C8}$$

Now define, following Minkowski (1903),

$$A = -\boldsymbol{n}\cdot\boldsymbol{r}_{\vartheta\vartheta}\,,\quad B = -\boldsymbol{n}\cdot\boldsymbol{r}_{\vartheta\psi}/\sin\vartheta\,,\quad C = -\boldsymbol{n}\cdot\boldsymbol{r}_{\psi\psi}/\sin^2\vartheta\,, \tag{C9}$$

so the Gaussian surface density can be expressed as

$$G(\vartheta,\psi) = AC - B^2\,. \tag{C10}$$

Using the relations $\boldsymbol{n}\cdot\boldsymbol{r}_\vartheta = 0$, $\boldsymbol{n}\cdot\boldsymbol{r}_\psi = 0$ and $(1/\sin\vartheta\,\boldsymbol{n}_\psi)_\vartheta = 0$, we obtain expressions for A, B and C by means of the support function:

$$A = (\boldsymbol{n}\cdot\boldsymbol{r})_{\vartheta\vartheta} - \boldsymbol{n}_{\vartheta\vartheta}\cdot\boldsymbol{r} = \varrho_{\vartheta\vartheta} + \varrho\,,$$

$$B = \left[\frac{1}{\sin\vartheta}(\boldsymbol{n}\cdot\boldsymbol{r})_\psi\right]_\vartheta = \frac{1}{\sin\vartheta}\,\varrho_{\vartheta\psi} - \frac{\cos\vartheta}{\sin^2\vartheta}\,\varrho_\psi\,, \tag{C11}$$

$$C = \frac{1}{\sin^2\vartheta}\,[(\boldsymbol{n}\cdot\boldsymbol{r})_{\psi\psi} - \boldsymbol{n}_{\psi\psi}\cdot\boldsymbol{r}] = \frac{1}{\sin^2\vartheta}\,\varrho_{\psi\psi} + \frac{\cos\vartheta}{\sin\vartheta}\,\varrho_\vartheta + \varrho\,,$$

since $\boldsymbol{n}_{\vartheta\vartheta} = -\boldsymbol{n}$ and $\boldsymbol{n}_{\psi\psi}\cdot\boldsymbol{r} = \cos\vartheta\, r_3 - \varrho$, the third component r_3 of $\boldsymbol{r}$ being obtained from Eq. (C5).

In principle, one could try to solve for the support function directly from Eq. (C10), a nonlinear second-order partial differential equation of Monge-Ampère-type, when the Gaussian surface density is known. There is, however, no practical procedure for doing so. One suitable indirect method for obtaining the support function is the minimization scheme introduced in Sect. 6.

References

Bowell E., Hapke B., Domingue D., et al., 1989, Application of photometric models to asteroids. In: Binzel R., Gehrels T., Matthews M. (eds.) Asteroids II, Univ. of Arizona Press, Tucson, p. 524

Cellino A., Zappalá V., Farinella P., 1989, Icarus 78, 298

Craig I. J. D., Brown J. C., 1986, Inverse Problems in Astronomy. Adam Hilger, Bristol

Gill P. E., Murray W., Wright M. H., 1981, Practical Optimization. Academic Press, New York

Horn B. K. P., 1984, Proceedings of the IEEE 72, 1671

Karttunen H., 1989, A&A 208, 314

Karttunen H., Bowell E., 1989, A&A 208, 320

Karttunen H., Muinonen K., 1991, A&A 242, 513

Little J. J., 1983, An iterative method for reconstructing convex polyhedra from extended Gaussian images. In: Proc. of the National Conf. on Artificial Intelligence. Washington, D.C., p. 247

Lumme K., Bowell E., Zellner B., 1980, Lunar and Planetary Science VI, 637

Lumme K., Bowell E., 1981a, AJ 86, 1694

Lumme K., Bowell E., 1981b, AJ 86, 1705

Lumme K., Karttunen H., Bowell E., 1990, A&A 229, 228

Magnusson P., Barucci M., Drummond J., et al., 1989, Determination of pole orientations and shapes of asteroids. In: Binzel R., Gehrels T., Matthews M. (eds.) Asteroids II, Univ. of Arizona Press, Tucson, p. 66

Menke W., 1984, Geophysical Data Analysis, Academic Press

Minkowski H., 1903, Math. Ann. 57, 447

Nirenberg L., 1953, Comm. Pure Appl. Math. 6, 337

Ostro S. J., Connelly R., 1984, Icarus 57, 443

Ostro S. J., Connelly R., 1988, Icarus 75, 30

Russell H. N., 1906, ApJ 24, 1

Weissbluth M., 1980, Atoms and Molecules. Academic Press, New York, Chapters 2 and 4