2

N71-21019

Reports of the Department of Geodetic Science
Report No. 147

NASA CR-117404

GRAVITY FIELD REFINEMENT BY SATELLITE TO SATELLITE DOPPLER TRACKING

by
Charles R. Schwarz

Prepared for
National Aeronautics and Space Administration
Washington, D. C.

Contract No. NGL 36-008-093
OSURF Project No. 2514

The Ohio State University
Research Foundation
Columbus, Ohio 43212

December, 1970

2

N71-21019

Reports of the Department of Geodetic Science
Report No. 147

NASA CR-117404

GRAVITY FIELD REFINEMENT BY SATELLITE TO SATELLITE DOPPLER TRACKING

by
Charles R. Schwarz

Prepared for
National Aeronautics and Space Administration
Washington, D. C.

Contract No. NGL 36-008-093
OSURF Project No. 2514

The Ohio State University
Research Foundation
Columbus, Ohio 43212

December, 1970

Reports of the Department of Geodetic Science

Report No. 147

GRAVITY FIELD REFINEMENT BY SATELLITE TO SATELLITE DOPPLER TRACKING

by

Charles R. Schwarz

Prepared for

National Aeronautics and Space Administration
Washington, D. C.

Contract No. NGL 36-008-093

OSURF Project No. 2514

The Ohio State University
Research Foundation
Columbus, Ohio 43212

December, 1970

PREFACE

This project is under the supervision of Ivan I. Mueller, Professor of the Department of Geodetic Science at The Ohio State University, and it is under the technical direction of Jerome D. Rosenberg, Project Manager, Geodetic Satellites Program, NASA Headquarters, Washington, D. C. The contract is administered by the Office of University Affairs, NASA, Washington, D. C. 20546.

ABSTRACT

Two concepts of satellite to satellite tracking are studied by means of simulated least squares solutions for parameters describing the gravity field. The first concept uses the range rate between two satellites near together in very low orbits. In the second concept, a constellation of very high geostationary satellites track a single satellite in a very low orbit.

The experimental results indicate that better resolution of the gravity field can be obtained from two very low satellites. However, satisfactory results can also be obtained when a high geostationary satellite tracks the low satellite. The latter concept is recommended, since it also offers several operational advantages. A single low satellite is shown to be sufficient, although more resolution might be provided by using several satellites at different inclinations. The amount of gravimetric detail that can be resolved depends directly on the altitude of the low satellite. The minimum feasible altitude is considered to be 200 km, and from this altitude features as small as squares 200 km on a side may be resolved.

Because of the loss of detail with altitude, satellite to satellite tracking cannot replace surface gravimetry for extremely detailed local surveys of areas smaller than 200 km squares. However, it can effectively fill the gap between the very detailed information obtained from surface gravimetry and the broad scale information obtained from conventional satellite gravimetry. Satellite to satellite Doppler tracking promises to refine our knowledge of the gravity field both by performing fairly detailed surveys of ocean areas and by surveying the gravity field on a global basis.

ACKNOWLEDGEMENTS

The material contained in this report was presented as a dissertation in partial fulfillment of the requirements for the degree Doctor of Philosophy at The Ohio State University. Professor Ivan I. Mueller, serving as the author's advisor, helped to define clearly the problem to be studied, and he guided the author past many distractions to the central purpose of finding what resolution of the gravity field may be obtained from satellite to satellite Doppler tracking. Professors Richard H. Rapp and Urho A. Uotila served on the author's reading committee, offering suggestions to help clarify several points. A large amount of the computer time necessary to carry out the numerical simulations described in this report was furnished by the Instruction and Research Computer Center of The Ohio State University.

The idea of investigating satellite to satellite tracking was first suggested to the author by Dr. Milo Wolff, then at the Instrumentation Laboratory of the Massachusetts Institute of Technology. Dr. Wolff kindly discussed some results of his investigations with the author before their publication. The author has also had several interesting discussions and a continuing correspondence on the subject with Mr. Foster Morrison of the Geodetic Research and Development Laboratory, National Oceanic and Atmospheric Administration.

A very thorough proofreading of both the text and the equations was performed by Mr. Georges Blaha, who suggested wording which helped clarify the meaning of several sentences. Finally, the author wishes to thank Mrs. Evelyn Rist and Miss Barbara Beer for their excellent typing of the manuscript.

CONTENTS

1. INTRODUCTION

Since the advent of the space age, scientists engaged in satellite gravimetry have produced a series of progressively more accurate and more detailed global models of the earth's gravity field. These models are usually published in the form of a list of coefficients (C_{nm}, S_{nm}) in the spherical harmonic expansion of the gravitational potential,

$$V = \frac{GM}{r} \sum_{n=0}^{\infty} \left(\frac{a}{r}\right)^n \sum_{m=0}^{n} \left(C_{nm} \cos m\lambda + S_{nm} \sin m\lambda\right) P_{nm} (\sin\varphi),$$

since this form is the most suitable for the computation of satellite motion. In this sense, one model gravity field is said to be more detailed than another if it contains tesseral coefficients of higher degree and order. The most detailed gravity field presently published is the SAO C 20.5 field, which constitutes a major part of the Smithsonian Astrophysical Observatory 1969 Standard Earth [Gaposchkin and Lambeck, 1970]. This field is complete through degree and order 16, with some isolated coefficients to degree 22. The Naval Weapons Laboratory has formed a larger gravity model containing 450 coefficients, although the coefficients of this field have not been published [Anderle, 1970]. This field, designated NWL 9B, is complete through degree and order 19, with some coefficients up to degree 26.

The steady improvement in the accuracy of modern gravity field models is due to improved tracking accuracies as well as the gathering of data from satellites of different inclinations. This improvement has been accompanied by a rise in the goals of satellite gravimetry. The most important need for gravity field models in the last decade has been for use in predicting satellite motion. Ideally, the parameters describing the gravity field should be sufficiently accurate that the error arising from this source in predicting the position of a satellite at a future epoch should be smaller than the certainty with which the position of the satellite at the future epoch can be measured. Although this goal still seems to be out of

reach, modern gravity field models can fulfill most practical satellite position prediction needs. Probably the most stringent requirements on satellite position prediction are imposed by the needs of satellite navigation. This application requires that the position of the satellite be immediately available to the navigator at the time he measures some function of his position relative to it, even though the orbital elements of the satellite may have been determined by semi-permanent ground based tracking stations some 24 to 48 hours previous. The process of updating the orbit from the epoch of its most previous determination to the epoch at which the satellite is observed by the navigator requires knowledge of the gravity field, or at least the low order zonal and tesseral coefficients.. Simulations by Anderle, et al. [1969] indicate that the effect of geopotential terms above the 12th degree and order on a satellite at a height of 600 nautical miles are generally less than 30 meters during a 24 hour period. As the time span increases or the altitude decreases, the prediction error caused by neglected terms will increase. Comparisons performed by Douglas and Marsh [1970] using GEOS-I and GEOS-II observations indicate that even with the best available gravity models, the satellite position is uncertain by 50-100 meters for heavily observed 5-6 day arcs. This is the precision with which a 5-6 day orbit can be fit to the data. The prediction capabilities of the various models investigated are considerably worse. The best fits of the observed data about orbits predicted from earlier observations were obtained with the SAO 1969 Standard Earth model, which gave rms errors of 100-150 meters for a 6 day prediction of GEOS-I and errors about three times as large for a five day prediction of GEOS-II. Comparisons reported by Wong and Prislin [1970] indicated substantially the same estimates for the precision of orbit fits to 6 day arcs using available gravity models.

A second requirement for accurate gravity models is imposed by the scientific goal of geodesy of determining the shape of the geoid. The simulations performed by Anderle, et. al. [1969] indicate that the geoid height may

be computed with an error of less than 10 meters from a gravity model truncated at degree and order 12. An analysis by Rapp [1967] indicates that the rms error in geoid height committed by neglecting terms in the geopotential of degree higher than 12 is about 4 meters. The geoid computed from the SAO 1969 Standard Earth gravity field parameters is estimated to be accurate to three meters in most areas [Gaposchkin and Lambeck, 1970], although it may be somewhat less reliable in areas where no surface gravity data were available [Gaposchkin, 1970].

Future requirements of satellite gravimetry are discussed in the report of the Williamstown Conference [Kaula, 1969]. Oceanographers will require a determination of the geoid to an accuracy of 10 cm, together with orbit determination capabilities allowing the computation of satellite altitude accurate to 10 cm. These accuracies will allow the measurement of the pelagic sea state with comparable accuracies by satellite altimetry. These requirements are at least two orders of magnitude beyond present capabilities. Furthermore, they appear to be beyond the capabilities of any land based satellite tracking system in the foreseeable future. It is expected that pulsed laser systems capable of tracking the range to a satellite with an accuracy of 10 or even 5 cm will ultimately be available. However, even this accuracy of satellite tracking will not allow the determination of terms in the harmonic expansion of the geopotential above degree 22 or so with the present satellites which are equipped with laser retroreflectors [Gaposchkin, 1970]. The reason is that the position of a satellite, at least at altitudes normally used for geodetic satellites, is just not sufficiently sensitive to the high order terms in the gravity field. This does not rule out the possibility that the velocity of a satellite might show significant variations of short time duration due to these high order terms. It is not unreasonable to expect that a satellite's velocity might show variations that cannot be determined by position measurements, in the same way that the slope of the geoid exhibits some rather large scale variations within small areas, while the geoid itself appears to be a fairly smooth surface, at least on a large scale.

Several combinations of satellite determinations of the gravity field with surface gravimetry have been published [Kaula, 1966a; Rapp, 1969]. Since the satellite data contributes most strongly to the low degree terms, the contribution of the surface gravimetry is most significant in the higher order terms. For instance, the satellite data used in computing the parameters of the SAO 1969 Standard Earth was sufficient to give a strong solution only for terms in the geopotential through degree and order 12, as well as some higher degree terms which produced resonant effects; the extension of the field to degree and order 16 was possible only by the inclusion of terrestrial gravity data [Gaposchkin and Lambeck, 1970; Gaposchkin, 1970]. On the other hand, the reliability of the potential coefficient information implied by the surface gravimetry is somewhat suspect because of the large areas that still remain gravimetrically unsurveyed. Therefore it is still desirable to measure the high order - short wave length components of the gravity field on a global basis by satellite methods, and to use the surface gravimetry to provide an independent check on the satellite determination.

If ground based tracking of satellite position will not be able to measure the high order terms of the geopotential, then the computation of gravity field models by this method must eventually end, and new methods of sensing the gravity field must be sought. Of the new methods currently being discussed, both satellite to satellite tracking and satellite to ocean altimetry can provide better refinement of the geopotential than can earth tracking of satellites [Lundquist, 1970].

Of these two methods, the concept of satellite to ocean altimetry has been developed in far more detail. The procedures by which such data might be used to determine the shape of the geoid have been considered by several investigators. Hudson [1970] describes a procedure in which the slope of the geoid is measured along the sub-satellite track by using the

rate of change of the measured altitude. This requires that the radial velocity of the satellite be independently known with an accuracy commensurate with the altitude measurement, which can be provided with present orbit determination methods and gravity models [Weiffenbach, 1969]. The concept described by Lundquist [1967] involves treating the altitude data as any other satellite tracking data and solving the geoid determination and orbit determination problems simultaneously. These procedures identify the mean sea level surface with the geoid. However, the mean physical sea surface is not an equipotential surface; stable departures of the mean sea surface from the geoid exist because of such factors as currents, variations in temperature, and variations in salinity. Whereas the height of the geoid above a mean earth ellipsoid can reach 100 meters, the maximum separation between the mean sea surface and the geoid is on the order of 10 meters [Kaula, 1969, p. 3-2]. In areas where sufficient measurements of temperature, pressure, and salinity exist, it may be possible to correct for up to 90% of this variation. Thus it is reasonable to identify the mean sea surface with the geoid if the accuracy of the altimeter measurement is on the order of one or a few meters, and so use the altimeter to map the geoid to this accuracy. However, oceanographers hope ultimately to use satellite altimetry to map the relief of the ocean with an accuracy of 10 cm. [Kaula, 1969, p.2-2, p. 3-2]. This will require that the geoid be known to the same accuracy, and that the geoid determination be independent of the ocean relief measurement. For this reason, the ability of satellite to satellite tracking to provide refinement of the gravity field deserves careful consideration.

1.1 The Concept of Satellite to Satellite Doppler Tracking

Although the use of one satellite to track another is a simple extension of the usual concept of tracking satellites from the ground, there has been very little discussion in the literature of the kind of information that might be obtained from such a system. The use of satellite to satellite tracking in orbit determination was mentioned a decade ago [Baker, 1960]. More

recently, the possibility that such tracking might be used to refine our knowledge of the gravity field has been discussed by Wolff [1969], and in [Kaula, 1969].

Although one satellite could conceivably track another by any of the means that have been used for ground based tracking of satellites, the most practical choice appears to be a two-way Doppler system. The necessary weight limitations rule out optical and pulsed laser systems since these systems require precise pointing. The Doppler equipment appears to be preferable to range measuring systems because it is both simpler and can be built to produce higher relative accuracy. Furthermore, small variations in velocity are more directly related to small variations in gravity than are variations in position. Therefore, the measurement of the range rate between two satellites by Doppler equipment is the only tracking mode considered in this study.

The concept proposed by Wolff [1969] employs two satellites in exactly the same circular orbit, with one following the other at a distance of 100-200 miles. Since they are near together, both satellites are affected in approximately the same manner by the low degree - broad scale components of the gravity field, so that the range rate between the two satellites is insensitive to these components. Conversely, the intersatellite range rate is most sensitive to features in the gravity field smaller than the intersatellite distance. Wolff proposes using the relative velocity of the two satellites as a direct measure of the difference in their kinetic energy, relating this in turn to a difference in potential energy and finally to a difference in gravitational potential. Since both satellites are assumed to follow precisely the same orbit, the difference in gravitational potential between the two satellites can be regarded as a difference in potential at different points on one of the orbits. Thus, the result of a single pass is a profile of values of the gravitational potential on a sphere whose

radius is that of the orbit. A series of such profiles will enable one to draw a contour map of the values of the potential on this sphere.

This concept was examined in some detail in the studies described in Chapter 2. These studies showed that although the observation that variations in the range rate of two satellites close together in the same orbit reflect small features in the gravity field while remaining insensitive to large scale features is substantially correct, other parts of the theory involve assumptions that cannot realistically be fulfilled. Specifically, the validity of several of the mathematical manipulations depends on the assumptions that both orbits are precisely the same and precisely circular, assumptions that cannot possibly be realized in practice.

The concept of satellite to satellite tracking contained in the Williamstown report [Kaula, 1969] is quite different. Instead of two satellites in the same orbit, this report envisions one satellite in a low orbit that can be tracked by any of a constellation of three satellites in high geostationary orbits. This report recognizes the extreme importance of putting the low satellite into as low an orbit as possible, since the effect of small features in the gravity field falls away rapidly with increasing altitude [Needham, 1970, p. 6]. Thus the concept includes an air drag sensing system and compensating thrusters. Such a system not only extends the lifetime of a low satellite, but also enables the satellite to follow an orbit unperturbed by air drag [Lange, DeBra, and Kaula, 1969]. The use of a drag free satellite thus immediately removes the problem of the solution for the parameters describing the gravity field being affected by unmodelled forces due to air drag and solar radiation pressure. On the other hand, the weight of the thruster fuel required to maintain a drag free orbit is considerable, especially at altitudes below 200 km. A graph depicting the trade off of required propellant with orbital altitude and eccentricity is shown in the Williamstown report [Kaula, 1969]. From this graph, a perigee altitude of 200 km was chosen as the lowest that could be reached with reasonable satellite lifetime. This minimum perigee altitude places

an effective bound on the resolution of the gravity field that can be obtained with satellite to satellite tracking.

The three geostationary satellites are spaced equidistant in longitude (120° apart), so that the low satellite can be tracked at any point in its orbit by at least one of the three. In addition to measuring the range rate to the low satellite, the geostationary satellites relay the range rate data to the ground. The use of geostationary orbits for the high satellites affords two important advantages over other configurations. Not only is at least one high satellite available to track the low satellite at any time, but they also may be tracked from the ground by permanently pointed antennas with high accuracy. The concept described in the Williamstown report includes monitoring of the position of the three geostationary satellites with permanently pointed laser and very long baseline interferometry equipment.

The main disadvantage of the system described above is that only the low satellite is significantly perturbed by variations in the gravity field. This means that the range rate measured by this system will contain the effect of large scale variations in the gravity field as well as the small scale variations that are of primary interest. While the range rate expected between two satellites slightly separated in the same orbit is only a fraction of a meter per second, the range rate between a high geostationary and a low minimum altitude satellite may be several thousand meters per second. Although the effects of the differences in orbital parameters and the effects of the low order terms in the geopotential can be separated mathematically, the use of geostationary satellites to track minimum altitude satellites provides a far less direct measurement of the fine structure of the gravity field than does the use of two satellites close together. Furthermore, if only one satellite is in a low orbit, the effects of fine detail in the gravity field on the range rate are accumulative, whereas if both satellites are in the same low orbit these effects are largely transitory. This means that the configuration of two satellites

in the same low orbit is more capable of discerning detail in the gravity field than is the configuration of one minimum altitude satellite tracked by geostationary satellites. This phenomenon is discussed in more detail in Chapter 5.

The major advantage of a satellite to satellite Doppler system is that the measurement can be made with far greater accuracy than the range rate can be measured by ground based stations. The main reason for this is that both satellites are above the troposphere, so that tropospheric refraction effects are eliminated. Ionospheric refraction effects are also somewhat lessened, although the radio link between a high geostationary satellite and a satellite at a minimum altitude of 200 km will traverse 90% of the ionosphere [Kaula, 1969, pp. 2-29]. Therefore, the determination of ionospheric refraction will still be necessary. This will be done by comparing the Doppler shift on two different frequencies, in the same manner as it is done for ground to satellite Doppler measurements [Weiffenbach, 1967]. The accuracy estimate for the satellite to satellite range rate measurement contained in the Williamstown report is 0.3 to 1.0 mm/sec with present technology, with an accuracy of 0.03 to 0.05 mm/sec eventually being possible [Kaula, 1969]. In most of the simulated solutions contained in this study, a standard deviation of 0.05 mm/sec is used to form the weights for the range rate equations.

Although the Williamstown report contains a great deal of discussion of the instrumentation that could be used in a satellite to satellite tracking system, very little is said about how the data might be reduced and analyzed. It is assumed that information concerning the gravity field of the earth can be extracted from the tracking data gathered by such a system for two reasons:

(1) Since the position of the geostationary satellite is constantly

monitored, the use of these satellites to track the minimum altitude satellite provides the same kind of information as tracking data taken at fixed known locations on the ground. Thus, given sufficient accuracy in all components of the system, ephemerides of the low satellite may be assembled and analyzed for the effects of irregularities in the gravity field.

(2) Scientists at the Jet Propulsion Laboratory have demonstrated an ability to determine features in the lunar gravity field by analyzing Doppler tracking data of lunar orbiters [Muller and Sjogren, 1968a and 1968b]. Since a very high satellite tracking a very low satellite presents an analogous situation, there is no reason why similar analysis should not resolve features of the earth's gravity field. In fact, the current interest in satellite to satellite doppler tracking appears to have been stimulated by the detection of the features in the lunar gravity field ascribed to "mascons".

The main purpose of this study is to answer the questions of how satellite to satellite range rate data might be reduced and analyzed, and what resolution of the gravity field might be obtained. As shown in the discussion above, there are several variable parameters to be considered in defining a satellite to satellite Doppler tracking system:

(1) Are better results obtained if a very high satellite tracks a very low satellite, or if both satellites are in low orbits? If both satellites are in low orbits, must they be in precisely the same orbit or is some variation in their relative configuration desirable? As previously mentioned, maintaining both satellites in precisely the same orbit is clearly impossible.

The discussion in Chapter 5 will show that some variation in the relative configuration of the two low satellites is desirable.

(2) What should be the altitude of the lower satellite? This depends on the amount of detail to be measured, and the minimum feasible altitude of 200 km discussed above places a limit on the amount of resolution that can be obtained.

(3) Is it necessary or desirable to have more than one low orbit? If so, are orbits of different inclinations necessary? Both of these questions are answered negatively.

(4) How accurate must the measurement of the range rate between the satellites be? Is it also necessary to track the low satellites from the ground? If so, how accurate must this tracking be? The assumed satellite to satellite range rate accuracy of 0.05 mm/sec has already been discussed. Ground tracking of the low satellites is necessary to provide some geographic location to the gravimetric phenomenon being observed, but the simulated experiments discussed in Chapter 5 show that highly precise tracking from the ground is neither necessary nor desirable.

(5) What data rate of range rate observations is desirable? The projected accuracy of 0.03 to 0.05 mm/sec contained in the Williamstown report is based on 10 second averaging of the Doppler signal, so that the data rate cannot be faster than one range rate measurement every ten seconds. The simulated solutions discussed in Chapter 5 show that a data rate of one range rate measurement every 30 seconds provides satisfactory results.

1.2 Gravity Field Representation

In order to discuss resolution of the gravity field in a specific manner, it is necessary to specify some method of representation of the gravity field. If the gravity field is represented by the spherical harmonic series for the geopotential, then obtaining more resolution means obtaining the coefficients of higher degree terms. However, there are several reasons why the spherical harmonic series is not considered appropriate to represent the detailed structure of the gravity field. For instance, each spherical harmonic coefficient is an integral over the total mass distribution of the earth. If some phenomenon of the gravity field, such as gravity anomalies or gravity disturbances, are observed, the spherical harmonic coefficient must be expressed as the integral over the whole earth involving this phenomenon. The measurement of range rate between two satellites is essentially a measurement of the gravity disturbance in the direction of the line joining the two spacecraft. If the gravity field is represented by the spherical harmonic series, then data taken over the whole earth must be included in each solution, and each solution is necessarily a global solution. Because the harmonic series representation is a global representation, it does not allow for local variations in our knowledge of the gravity field; i.e., it is not capable of reflecting the fact that we may know some phenomenon of gravity much better in some areas than in other areas. This indicates that it would be more appropriate to represent the gravity field by some kind of <u>local</u> representation such as the value of some phenomenon of the anomalous gravity field in some conveniently sized blocks on the surface of the earth or on the geoid. The main advantage of a local representation is that the blocks in some areas can easily be made smaller to reflect the greater amount of detail available for those areas.

The most familiar phenomenon to use is the gravity anomaly, since this is commonly used by geodesists and geophysicists. The gravity anomaly is the most appropriate quantity if the final purpose is the computation of geoid undulations according to Stokes' Formula. On the other hand, the

use of gravity anomalies for computing satellite orbits is quite cumbersome, since it is necessary to compute the derivatives of Stokes' function in each block at each step of the integration [Obenson, 1970]. However, despite the cumbersomeness of the formulae, the computer time required to compute an orbit from a spherical harmonic series containing a given number of coefficients is about the same as the computing time required when the accelerations are computed from the same number of mean gravity anomalies [Kaula, et al., 1966, p. III. D.-1-]. The use of a fictitious surface layer has been proposed by Koch [1968] as an alternate representation of the earth's gravity field. If the anomalous gravity field is represented by a fictitious surface layer whose surface density produces the disturbing potential, then the formulae for the disturbing acceleration acting on a satellite are far simpler in form than those in the case of gravity anomalies. Thus, the representation of the gravity field by the density of a fictitious surface layer may represent the best compromise between the needs of the geophysicist and the needs of the orbit analyst.

Another argument against using the spherical harmonic series to represent the gravity field in detail is that the number of coefficients required is just too great to be practicably feasible. The number of coefficients contained in a spherical harmonic series complete through degree and order (n, n) is $(n + 1)^2$. A simple computation shows that the number of equal area blocks required to cover the earth is also about $(n+1)^2$ if the side length of each block is 180/n degrees of arc. Since the shortest wavelength contained in the spherical harmonic series is 360/n degrees, it is reasonable to say that a block representation represents approximately the same amount of detail as a spherical harmonic representation if the side length of the blocks is equal to the half-wavelength of the highest degree term in the spherical harmonic series. This does not imply that the two representations are mathematically equivalent by any means, but it is intuitively obvious that if the gravity field has a strong component of wavelength 360/n degrees, this component will be well represented by blocks whose

side length is 180/n degrees.

The advantage of using a block representation rather than a spherical harmonic series in which the number of coefficients is approximately the same as the number of blocks is that it may not be necessary to consider all the blocks. If the satellite altitude is not too high, it may be possible to compute the anomalous acceleration acting on the satellite by considering the parameters describing the anomalous gravity field in only a few blocks in the vicinity of the sub-satellite point. Ideally, we would like to have a representation of the anomalous gravity field such that the disturbing potential at any altitude depends only on the parameter describing the block directly under the satellite. No known set of functions is able to afford complete directness of representation, although some quantities afford more than others. A comparison of the directness of representation afforded by representing the anomalous gravity field by mean gravity anomalies and that afforded by mean surface densities is discussed in Chapter 3. In the computation of the disturbing potential or the gravity disturbance at satellite altitude, the directness of the representation by mean surface densities is slightly greater than the directness of the representation by mean gravity anomalies, which is another way of saying that the influence of distant zones is smaller in the case of mean surface densities [Heiskanen and Moritz, 1967, p. 242].

If a method of representation of the gravity field is fairly direct, then the anomalous acceleration acting on a satellite depends on only a few blocks or a few functions; conversely, the anomalous acceleration experienced by the satellite when passing over an area can then be used to solve for the parameters describing the gravity field in this area independently of other areas. This means that if a fairly direct method of representation is used, every solution for parameters describing the gravity field need not be a global solution. If the motion of the satellite is observed only when the satellite is over certain areas, then it is possible to solve for

the parameters describing the gravity field in those areas without the solution being too greatly influenced by the neglect of the unmodelled effects of the gravity field in other areas. Thus, if the gravity field is of especial geophysical interest in some area, such as an ocean trench, it should be possible to examine the structure of the gravity field in that area by heavily observing the satellite as it passes over the area.

The importance of the ability to solve for only some parameters at a time, somewhat independently of other parameters, cannot be stressed too strongly. The description of the global gravity field in $1^{\circ} \times 1^{\circ}$ equal area blocks requires over 40,000 parameters, and no scientist has the resources to perform simultaneous solutions for this many unknowns as a matter of routine. From this viewpoint, the spherical harmonic series is the worst possible representation of the gravity field; every solution for the coefficients must include data gathered on a global basis and must include all the coefficients as unknowns. The description of the anomalous gravity field by the mean density of a surface layer in blocks, or by mass concentrations distributed in a regular grid, are probably the best from this viewpoint, since these methods provide more directness of representation than others. An attempt to describe the gravity field by a new method using a novel set of functions and affording great directness of representation has recently been described by Lundquist, et al. [1970].

The fictitious surface layer was selected to represent the gravity field for the simulations described in Chapter 5. The properties of such a representation are examined in more detail in Chapter 3, and an algorithm incorporating the concept of a fictitious surface layer is described in Chapter 4.

2. THE RANGE RATE BETWEEN TWO SATELLITES CLOSE TOGETHER IN LOW ORBITS

If two satellites are put into the same circular orbit in a purely central force field, the distance between them will remain constant and the rate of change of this distance will remain always zero. As soon as the orbit is changed from circularity, or as soon as the satellites are perturbed by a non-central force, the range rate between the satellites will depart from zero. However, the behavior of this range rate is not well known.

To study the behavior of the range rate, an extremely simplified situation was simulated. The force field of the earth was represented as a dominant central force plus the attraction of a single point mass placed on the surface of the earth in the plane of the equator. The mass assigned to this point mass was 10^{-6} earth masses. A single orbit at an altitude of 1700 km was numerically integrated in this force field. The initial conditions of this orbit were chosen so that the orbit would be perfectly circular in the absence of perturbing mass. Since both the orbit and the disturbing mass lay in the equator, the situation could be viewed in two dimensions. Two satellites were assumed to be traveling in this orbit, the second passing a given point 25 seconds after the first, and the range rate between the two satellites was computed. The constant time delay of 25 seconds corresponds to a linear separation of about 175 km.

The range rate between the two satellites for slightly over one revolution is shown in Figure 2.1. Two components of the range rate may be discerned. First, there is a periodic component whose period coincides with that of the orbit. This component is indicated by the dashed line. The amplitude of this component appears to increase secularly. Secondly, superimposed on the first component is a pattern which only appears when the satellite passes over the disturbing mass. This pattern can be seen

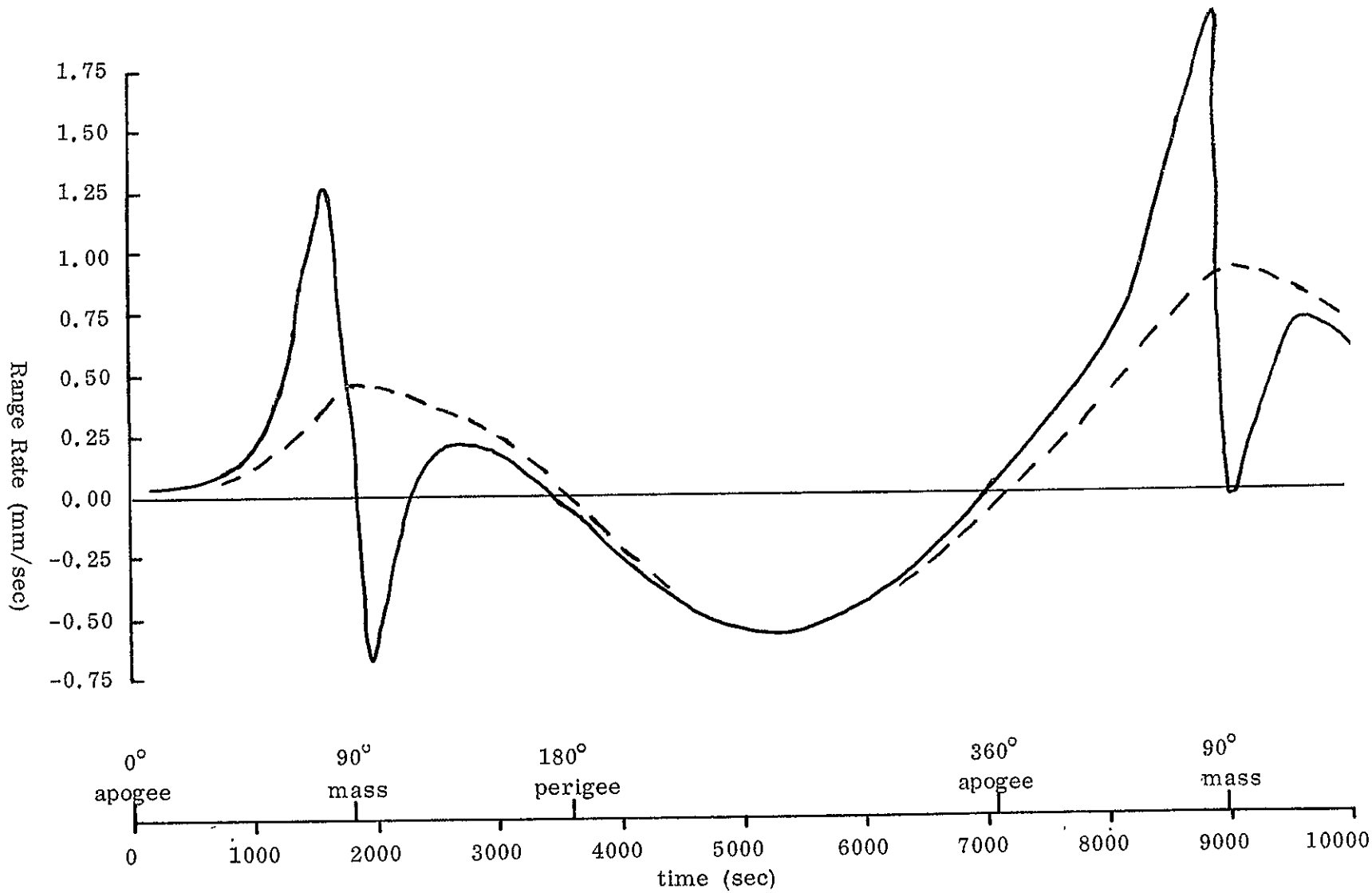

Fig. 2.1. Range Rate Between Two Satellites in the Same Orbit. Dashed Line Indicates Periodic Component.

clearly when the first component is subtracted from the actual range rate (Figure 2.2). It consists of an accelerating rise from zero to a sharp peak, followed by a fast drop to a negative extremum, followed by a return to zero. Since this pattern characterizes the range rate during the period the satellites are passing over the point mass, it may be called the characteristic signature of a point mass. Characteristic patterns for point masses and plate shaped masses are discussed by Kane [1969] for the case of a lunar orbiter tracked by Doppler equipment from the earth.

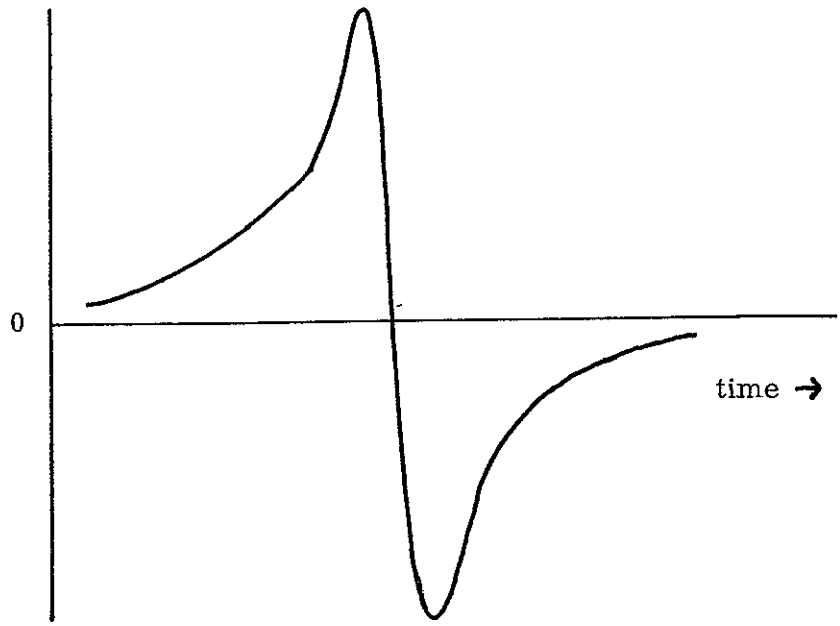

Fig. 2.2. Characteristic Signature of a Point Mass in the Range Rate Between Two Satellites

The characteristic signature shown in Figure 2.2 may be given an explanation that appeals to intuition by considering only the in track component of the disturbing force. As the two satellites approach the point mass with zero relative velocity, both are attracted and the velocity of both increases. However, the first satellite, being nearer to the

attracting source, is attracted more strongly. Its velocity increases faster than that of the second satellite, causing a net positive range rate as the distance between the satellites increases. The difference in the in track components of the attraction becomes more pronounced as the satellites approach the source, causing an acceleration in the graph of the range rate. When the two satellites approach within a few hundred kilometers of the attracting source, the in track components of attraction rapidly become equal again, so that the range rate reaches a maximum and ceases to increase. When the first satellite is directly over the attracting source, the horizontal component of the force with which it is attracted goes to zero; the forward velocity of the second satellite is still being increased, so the range rate between them is decreased. After the first satellite has passed the attracting source, its forward motion is retarded; the second satellite is still being attracted forward, the difference of the in track accelerations is sharply negative, and the net range rate rapidly falls to zero and becomes negative. After the second satellite has passed several hundred kilometers beyond the attracting mass, the in track accelerations again become equal and the range rate is at a negative extremum. From that point on, the second satellite, being nearer to the attracting mass, is more strongly retarded, so that the range rate tends to increase, eventually returning to zero.

The radial component of the force exerted by the attracting mass must also be considered. The effect of this force is to pull both satellites downward, increasing their velocity and increasing the eccentricity of the orbit. Even if the initial conditions are selected so that the orbit is initially circular, the disturbing mass will cause the orbit to become eccentric. It is impossible to maintain a precisely circular orbit in the presence of disturbing forces, so that eccentricity of the orbit must be expected. The eccentricity of the orbit used to generate the range rate shown in Figure 2.1 was initially zero, but after one revolution, it had increased to 0.000004.

The increase in the eccentricity may have both long period and secular terms.

The periodic component of the range rate, on which the characteristic signature is superimposed, is caused by the eccentricity of the orbit. The growth in the amplitude of this component reflects the increasing eccentricity. Since the angular and linear velocities of the satellites are greater at perigee than at apogee, the constant time difference of 25 seconds must correspond to a larger linear separation at perigee than at apogee. This means that the separation of the two satellites must increase from apogee to perigee, as shown by a positive range rate. Similarly, the negative range rate from perigee to apogee indicates a decrease in the distance between the two satellites.

Since the total energy is constant along the orbit, the difference in gravitational potential at the positions of the two satellites is the negative of the difference in their kinetic energies. Within the small range of velocities considered, the kinetic energy difference is linearly related to the linear velocity difference, which is very nearly the range rate between the two satellites. The difference in gravitational potential at the positions of the two satellites is shown (with the sign changed) in Figure 2.3. Comparison with Figure 2.1 shows that the difference in gravitational potential is directly proportional to the range rate.

The analysis by Wolff [1969] suggests that the range rate is also directly proportional to the rate of change of gravitational potential along the orbit, so that the actual potential may be obtained (except for a constant of integration) by integrating the range rate along the orbit. In this highly simplified example, this relationship is very nearly true. The actual gravitational potential along the orbit is shown in Figure 2.4. The slope of this graph is very nearly directly proportional to the potential difference in Figure 2.3 or to the range rate in Figure 2.1. The dominant component

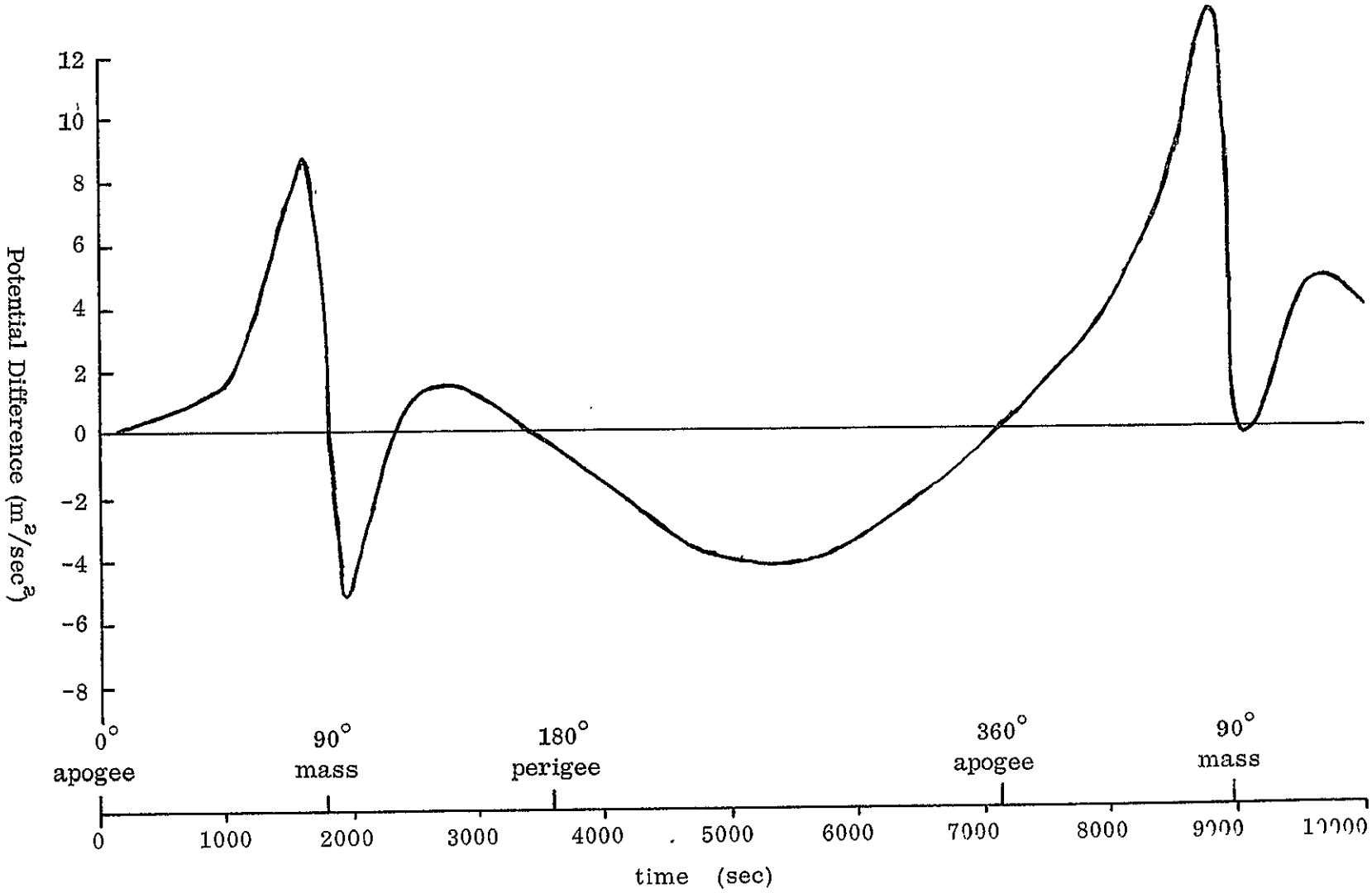

Fig. 2.3. Difference of Gravitational Potential for Two Satellites in the Same Orbit

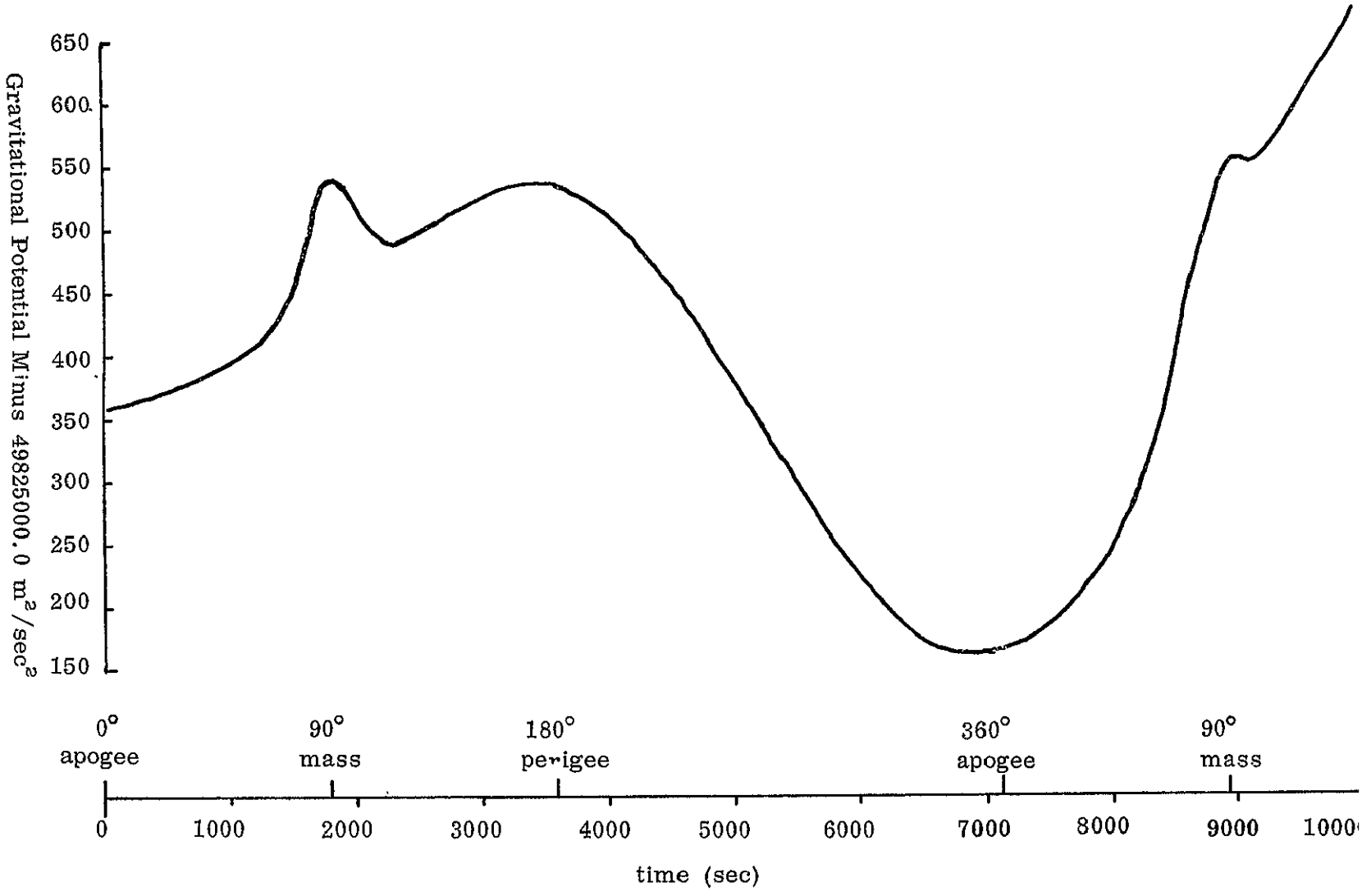

Fig. 2.4. Gravitational Potential Along the Orbit Perturbed by a Single Point Mass

in the graph of the potential is caused by the eccentricity of the orbit, as evidenced by the minimum at apogee and the maximum at perigee. The increasing amplitude of this component reflects the increasing eccentricity. The presence of the point mass is evidenced by a small "bump" in the graph, indicating a "bump" in the gravity field.

The component due to the eccentricity may be identified by its period and removed. The remaining component contains the "bump" and approximates a profile of the gravitational potential along an arc of a circle whose radius is the mean radius of the orbit.

Similar simulations were performed for several more complicated situations, again using two satellites in exactly the same orbit but using several point masses. The range rate computed for an orbit perturbed by three point mass is shown in Figure 2.5. The periodic component due to the eccentricity is more difficult to identify than in the case of a single perturbing mass, although the signatures of the point masses are still quite evident. In this case, the range rate was again found to be directly proportional to the derivative of the gravitational potential along the orbit. The signatures of the point masses again correspond to "bumps" in the potential field, so that a circular profile of the gravitational potential may again be formed.

The relationships discussed above suggest the possibility of using many profiles to construct a contour map of the gravitational potential on a large sphere. However, these relationships appear to break down when the two satellites are in orbits that are almost, but not exactly, identical. Figure 2.6 shows the range rate between two satellies in slightly different orbits in the presence of 3 point masses. The two orbits are chosen such that the two osculating orbits are perfectly circular at the initial epoch, and the satellites are initially separated by 200 km. By the time the second satellite reaches the position occupied by the first satellite at the initial epoch, its elements have been slightly perturbed, so that the orbits are very

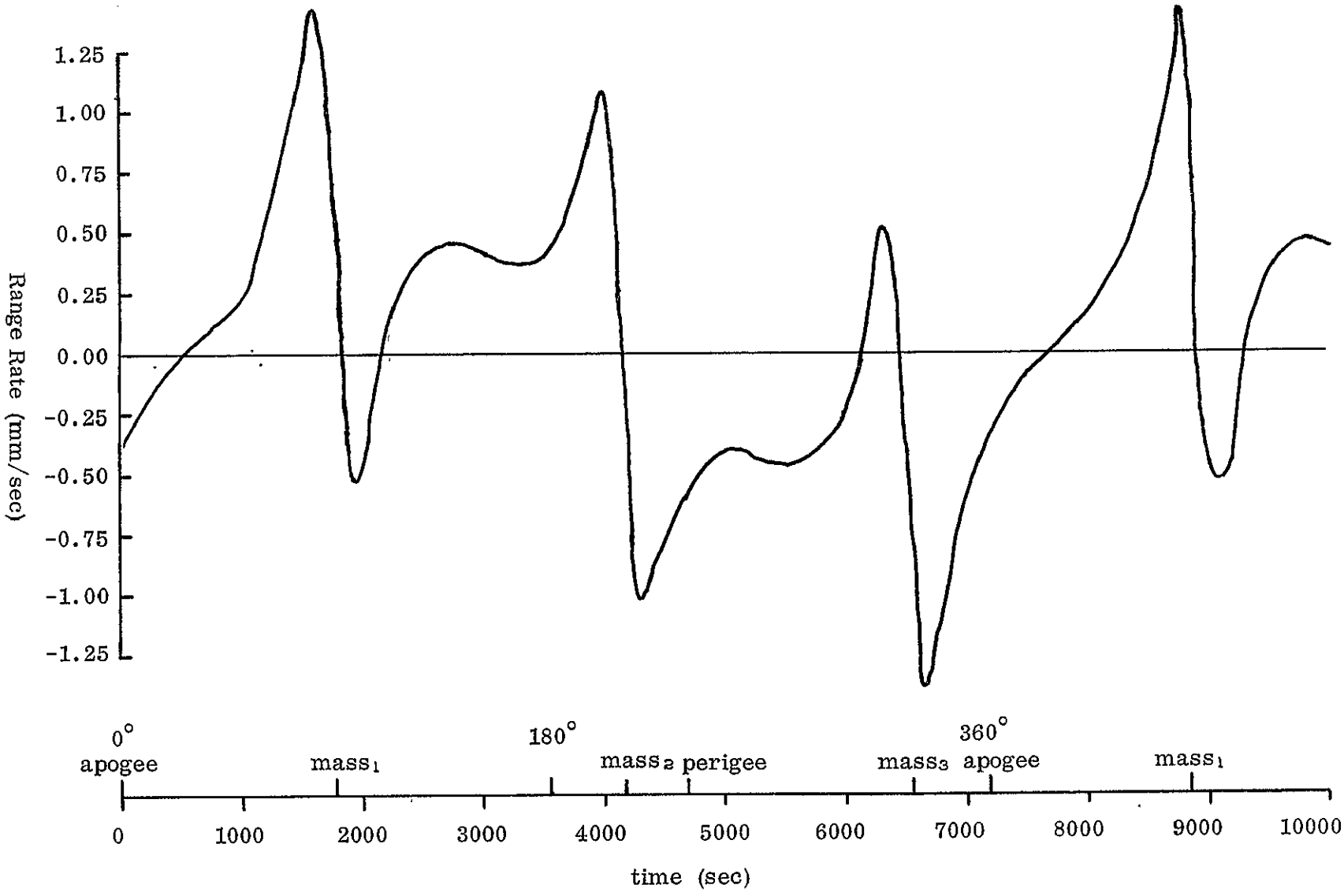

Fig. 2.5. Range RateBetween Two Satellites in the Same Orbit Perturbed by 3 Point Masses

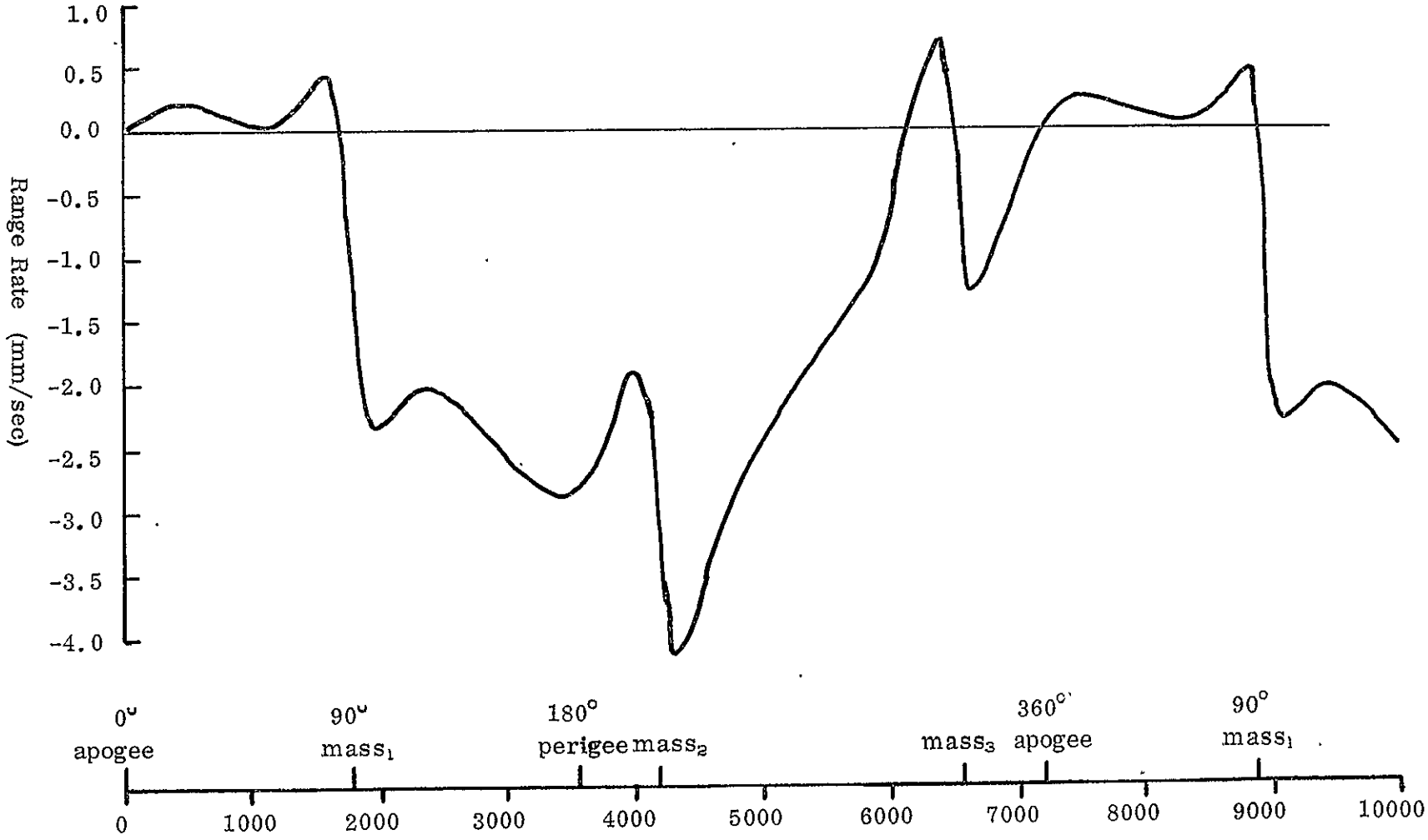

Fig. 2.6. Range Rate Between Two Satellites in Almost Identical Orbits

slightly different. The characteristic signatures of the three masses are still evident in Figure 2.6, although somewhat distorted. The periodic component caused by the eccentricity of the orbit is difficult to discern. The difference of the gravitational potentials at the respective locations of the two satellites is shown in Figure 2.7. The horizontal line is drawn at $-3.3\ m^2/sec^2$, which is the difference of the energy constants of the two orbits.

Although Figure 2.7 resembles Figure 2.6, the two graphs are clearly not scaled versions of one another. The relation between the linear velocities and kinetic energy still holds, so that the difference in gravitational potential in Figure 2.7 is a scaled version of a graph of the difference between the linear velocities of the two satellites (not shown). However, measurement of the range rate cannot adequately produce the difference in kinetic energy. The reason for this may be easily explained. The vector velocity of each satellite may be resolved into components in the direction of, and perpendicular to, the line joining the two spacecraft. If the perpendicular components are not nearly equal, then a significant amount of the difference in linear velocity is not measured by the range rate. Detailed analysis of the velocity vectors showed this to be the case for these two simulated orbits.

The relation between the difference in the gravitational potentials and the actual potential along the orbit also breaks down when the two orbits are not precisely the same. Figure 2.8 shows the gravitational potential along the path of the first satellite in this example. The graph for the second satellite (not shown) is quite similar. The "bumps" in the potential field caused by the point masses are quite evident, and the effect of the eccentricity can be readily discerned. However, the function shown in Figure 2.8 is clearly not the integral of the function shown in Figure 2.7. Several points of discrepancy may be discerned; the most noticeable is the span from 2000 to 4000 seconds, where the potential difference is negative

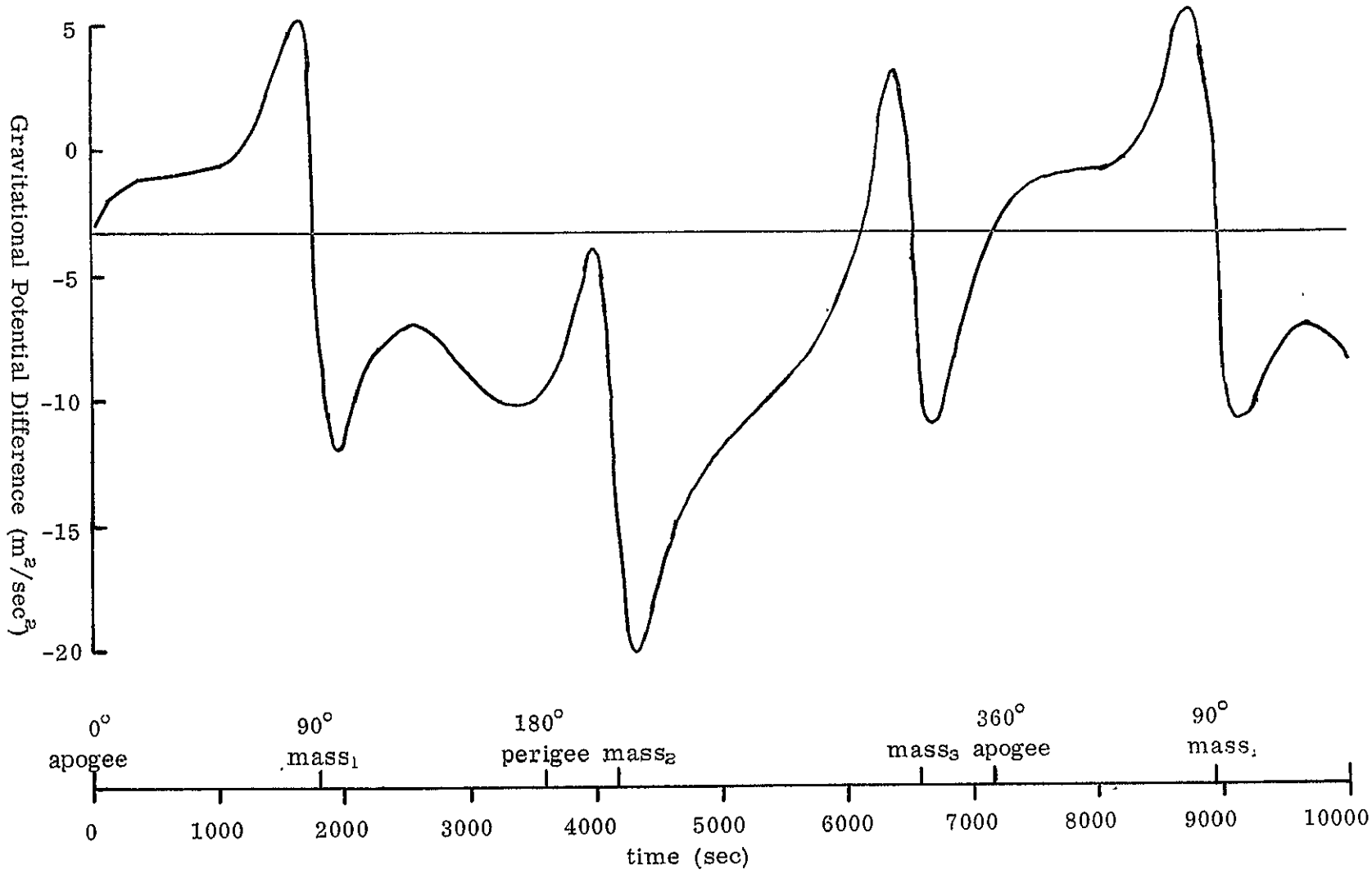

Fig. 2.7. Difference in Gravitational Potential for Two Satellites in Almost Identical Orbits

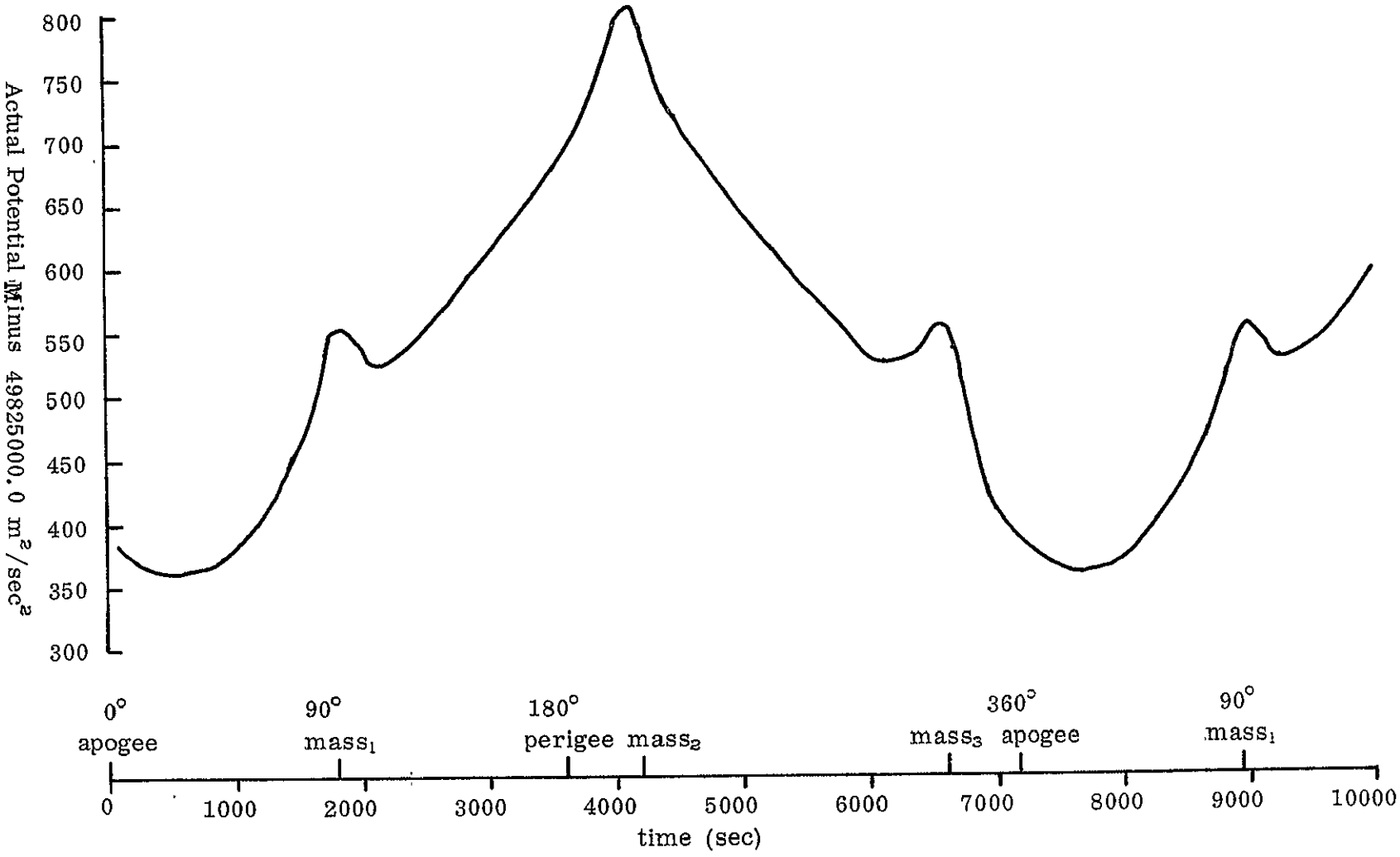

Fig. 2.8. Gravitational Potential Along the Path of the First Satellite

but the actual gravitational potential is increasing. Thus, the relationship that held for two satellites in precisely the same orbits no longer holds when the orbits are slightly different.

The example above shows that although a simple relationship exists between the potential and the range rate when both satellites are in precisely the same orbit, this relationship breaks down in two places when the orbits differ slightly. It is reasonable to expect that the relationship would break down still farther if a more complicated potential field, such as that of the actual earth, were considered. Since two satellites cannot be kept in precisely the same orbit, it is not reasonable to expect that the range rate between two orbiting satellites can be used to map the actual potential field of the earth directly onto a sphere.

3. THE POTENTIAL OF A FICTITIOUS SURFACE LAYER

Through the analysis discussed in the previous chapter, it was seen that satellite to satellite range rate does not provide a direct measure of the potential at the satellite position. Therefore, the possibility of mapping the gravitational potential function on a sphere by measuring profiles of its values was discarded. Since the measurement does not lend itself to the direct production of a contour map of the potential, a mathematical representation was felt to be more appropriate. A mathematical representation, either by a spherical harmonic series or by the values of some function of gravity in some convenient sized blocks, also lends itself to statistical analysis, which is quite difficult for functions represented empirically by contour maps. The number of parameters used to describe the gravity field mathematically depends on the detail that can be resolved. On the other hand, a satellite to satellite Doppler system will provide for more measurements of range rate than would be needed for a unique solution. Therefore, the classical method of least squares, long the favored approach of geodesists, was deemed to be the most appropriate method of data reduction and analysis.

Using this method, it is necessary to write the observed quantity, the range rate between the satellites, in terms of the unknowns of the problem, which in this case would be the orbital elements of each satellite together with the set of unknowns describing the gravity field. This relationship is then linearized around approximate values of the unknowns to provide a linear observation equation. An algorithm for forming and solving these equations is described in the next chapter.

For the reasons discussed in Chapter 1, it was felt that a simulation of a solution for a global description of the gravity field is not necessary. The huge number of unknowns involved in global solutions make such solutions impractical.

Neither were sufficient computer resources available to this investigator to simulate a series of global solutions. Furthermore, should range-rate data between satellites become available, the analysis of the data will involve using the data taken over a localized area to solve for the parameters describing the gravity field in that area. Thus, it was felt that a local representation of the gravity field by some phenomenon in blocks should be used, and solutions should only be simulated for localized areas containing a reasonable number of blocks.

The parameter chosen to represent the anomalous gravity field was the mean density of a fictitious surface layer. Although the use of gravity anomalies in blocks might be a preferable representation for geodesists, it was felt that the obtainable resolution of the gravity field could be equally well demonstrated with either method. The fictitious surface layer representation was chosen because it affords slightly more directness of representation and much simpler formulae for orbit integration than does the gravity anomaly representation. The fictitious surface layer is also of some geophysical interest, since it is likely that the small features of the gravity field are caused by anomalous mass distribution in the crust. Thus, a pronounced feature of limited extent in the fictitious surface layer will probably indicate a real excess or deficiency of mass near to the physical surface.

3.1 The Density of a Fictitious Surface Layer

As is conventional in physical geodesy, the total geopotential function is split into a normal potential and a disturbing potential,

$$W = U + T.$$

The function U is the potential of a level ellipsoid which contains the same mass as the earth M, has the same second order spherical harmonic coefficient J_2, has the same potential as the geoid W_o, and rotates with the

same angular velocity as the actual earth ω. This function can be mathematically described either in closed form or by a series of even degree zonal spherical harmonics whose coefficients are completely determined by M, J_2, W_o, and ω. The disturbing potential T can be represented as a series of spherical harmonics, as a function of gravity anomalies on the geoid, or as the potential of a fictitious surface layer, etc.

If $\varkappa$ is the density of a ficititious layer spread on the surface S, then the potential of this layer is

$$T = \iint_S \frac{G\varkappa}{\ell} \, dS \tag{3.1}$$

where G is the gravitational constant and ℓ is the distance from the point where T is evaluated to the integration element dS. It is convenient to let the parameter which describes the surface density be denoted φ, where $\varphi = G\varkappa$. Then φ has the same units as acceleration and gravity, and it is convenient to state its value in milligals. Thus, a value of φ of one milligal corresponds to an actual surface density of about 1.5×10^4 grams/cm^2. The expected order of magnitude is about the same as that of the gravity anomalies Δg. This may be seen at once from the formula [Heiskanen and Moritz, 1967, p. 303]

$$\varphi = \frac{1}{2\pi} \left(\Delta g + \frac{3}{2} \frac{\bar{\gamma}}{R} N \right) \tag{3.2}$$

where $\bar{\gamma}$ here denotes a mean value of normal gravity (979.8 gals), R is a mean radius of the earth (6371 km), and N is the geoid undulation. Since N/R is of the same order of magnitude as $\Delta g/\bar{\gamma}$, the second term in parentheses is of the same order as the first, and φ is of the same order as Δg. In fact, in many areas the second term in parentheses is somewhat smaller than the first term, so that the expression $\Delta g/2\pi$ is a fair approximation to φ. Since the geoid is a rather smooth surface, at least

when viewed in large areas, and Δg is a very rough function showing large and rapid fluctuations, a contour map of the function φ will contain the same pronounced features as a contour map of Δg.

A better approximation of the range of values of φ to be expected can be found by computing the variance of φ according to the method described in [Heiskanen and Moritz, 1967, Chapter 7]. Using the Stokes' Equation for the geoid undulation,

$$N = \frac{R}{4\pi\bar{\gamma}} \iint_{\sigma} \Delta g \; S(\psi) \, d\sigma ,$$

the surface density is expressed in terms of gravity anomalies alone as

$$2\pi\varphi = \Delta g + \frac{3}{8\pi} \iint_{\sigma} \Delta g \; S(\psi) \, d\sigma . \qquad (3.3)$$

Here σ denotes integration over the unit sphere and $S(\psi)$ is the Stokes' Function. Expressing the Stokes' Function in spherical harmonics [Heiskanen and Moritz, 1967, p.97] leads to

$$2\pi\varphi = \Delta g + \frac{3}{8\pi} \sum_{n=2}^{\infty} \frac{2n+1}{n-1} \int_0^{\pi} \int_0^{2\pi} \Delta g(\psi,\alpha) \; P_n(\cos\psi) \sin\psi \, d\alpha \, d\psi$$

$$= \Delta g + \frac{3}{2} \sum_{n=2}^{\infty} \frac{\Delta g_n}{n-1}$$

where Δg_n is the nth degree term in the spherical harmonic expansion of Δg, i.e., $\Delta g = \Sigma \, \Delta g_n$. Thus,

$$2\pi\varphi = \sum_{n=2}^{\infty} \left(1 + \frac{3}{2(n-1)} \right) \Delta g_n = \sum_{n=2}^{\infty} \frac{2n+1}{2(n-1)} \Delta g_n$$

The variance of φ is defined to be

$$\text{Var}\{\varphi\} = M\{\varphi^2\}$$

$$= \frac{1}{4\pi^2} M\left\{\sum_{n=2}^{\infty}\left(\frac{2n+1}{2(n-1)}\right)\Delta g_n \sum_{n'=2}^{\infty}\left(\frac{2n'+1}{2(n'-1)}\right)\Delta g_{n'}\right\}$$

$$= \frac{1}{4\pi^2}\sum_{n=2}^{\infty}\sum_{n'=2}^{\infty}\left(\frac{2n+1}{2(n-1)}\right)\left(\frac{2n'+1}{2(n'-1)}\right) M\{\Delta g_n \Delta g_{n'}\}$$

where the operator M denotes the mean value taken over the entire earth. Because of the orthogonality of the spherical harmonics,

$$M\{\Delta g_n \Delta g_{n'}\} = \begin{cases} c_n & \text{if } n' = n \\ 0 & \text{if } n' \neq n \end{cases}$$

where c_n is the nth degree variance of the gravity anomalies. Thus

$$\text{Var}\{\varphi\} = \frac{1}{4\pi^2}\sum_{n=2}^{\infty}\left(\frac{2n+1}{2n-2}\right)^2 c_n \,. \qquad (3.4)$$

The difficulty in evaluating this expression is that the degree variances c_n of the gravity anomalies are not well known. Although lists of degree variances of the gravity anomalies have been published by several investigators, the agreement between the different lists is not good. Furthermore, the published lists indicate that the degree variances decay quite slowly with increasing degree. Using data from Kaula, Pellinen [1970] has found that the degree variances for degrees 3 through 15 are adequately represented by the formula

$$c_n = .120\, n^{-1} \text{ mgal}^2$$

This formula cannot be valid for all n; if it were, the expression

$$\sum_{n=2}^{\infty} c_n = \text{Var}\{\Delta g\}$$

would not converge, implying that the rms point value of the gravity anomalies is not finite. However, it is reasonable to assume that the actual decay of the degree variances is represented by a formula of the form

where α is slightly greater than unity. If such a decay law is valid, then the series in (3.4) converges, although very slowly, and several hundred terms may be needed in order to evaluate it to even two significant digits.

Equation (3.4) may be approximately evaluated by using known values of the degree variances for low degrees and Pellinen's formula for high degrees. Thus,

$$\sum_{n=2}^{\infty} \left(\frac{2n+1}{2n-2}\right)^2 c_n = \sum_{n=2}^{N} \left(\frac{2n+1}{2n-2}\right)^2 c_n + \sum_{N+1}^{\infty} \left(\frac{2n+1}{2n-2}\right)^2 A n^{-\alpha}$$

The second sum may be approximated by

$$\int_{N+1}^{\infty} \left(\frac{2n+1}{2n-2}\right)^2 A n^{-\alpha} dn$$

$$= \frac{A}{(N+1)^{\alpha}} \left[\frac{N+1}{\alpha-1} + \frac{9}{4} \frac{N+1}{N} + \left(3 - \frac{9}{4}\alpha\right) \sum_{k=0}^{\infty} \frac{1}{(\alpha+k)(N+1)^k} \right]$$

for $\alpha > 1$. A set of degree variances of the gravity anomalies through degree 16 corresponding to the 1969 SAO Standard Earth gravity field is given by Gaposchkin and Lambeck [1970, p.72]. Using these values

$$\sum_{n=2}^{16} \left(\frac{2n+1}{2n-2}\right)^2 c_n = 402 \text{ mgal}^2$$

For the approximation of the higher degree variances by the formula $c_n = An^{-\alpha}$, Pellinen [1970] obtained the parameters A = 227, α = 1.15 from an analysis of the autocorrelation of actual gravity data. Using these values

$$\sum_{n=17}^{\infty} \left(\frac{2n+1}{2n-2}\right) c_n \approx 1013 \quad \text{mgal}^2$$

Substituting these values into equation (3.4) yields

$$\text{Var}\{\varphi\} \approx 35.8 \quad \text{mgal}^2 ,$$

so that

$$\text{rms}\{\varphi\} \approx 6.00 \text{ mgal.}$$

This is the rms of point values of φ. The rms mean density parameter for a block can be expected to be somewhat smaller.

The rms $\{\overline{\Delta g}\}$ for mean gravity anomalies in various size blocks are given by Moritz [1963]. The simple formula

$$\text{rms}\{\bar{\varphi}\} \approx \frac{1}{2\pi} \text{ rms}\{\overline{\Delta g}\}$$

may thus be used to gain an idea of the expected values of φ for various sized blocks. Such a set of values is given below.

Block Size	rms $\{\bar{\varphi}\}$
1° × 1°	4.6 mgal
2° × 2°	3.9
5° × 5°	3.1
10° × 10°	2.6

3.2 Directness of Representation of a Fictitious Surface Layer.

The property of directness of a representation of the anomalous gravity field means that the anomalous acceleration experienced by a satellite is fairly directly related to the parameters describing the gravity field in the few blocks in the vicinity of the sub-satellite point. If a representation has this property, then the anomalous velocity of each satellite and the range-rate between the satellites are also directly related to these parameters. More important, the dependence of the range rate on the parameters describing far away blocks will be negligibly small, so that these parameters may be neglected in the mathematical model relating the range rate to the gravity field. Thus, directness of representation is the same as the problem of the influence of distant zones in the upward continuation of gravity disturbances. Table 6-1 in [Heiskanen and Moritz, 1967] shows that the influence of distant zones is somewhat less in the case of a surface layer representation by gravity anomalies. The influence of distant zones on the disturbing potential in the cases of these two representations may also be compared. If we represent the disturbing gravity field by mean gravity anomalies in blocks, the disturbing potential is

$$T = \sum_k T_k = \frac{R}{4\pi} \sum_k \Delta g_k \, S(r, \psi_k) \, \Delta\sigma_k$$

where $S(r, \psi)$ is the extended Stokes' function and $\Delta\sigma_k$ denotes the area of block k in solid angle. If we use a mean value of R = 6371 km, then an anomaly of one milligal in a block of unit area will contribute 5.07 $S(r, \psi_k)$

m^2/sec^2 to the disturbing potential. Similarly, if we use the same blocks but represent the gravity field by a surface layer, the disturbing potential is given by

$$T = R^2 \sum_k \frac{\varphi_k}{\ell(r, \psi_k)} \Delta\sigma .$$

Since $\varphi \approx \frac{1}{2\pi} \Delta g$ as a crude approximation it is reasonable to compare a block in which $\varphi = \frac{1}{2\pi}$ mgal to a block in which $\Delta g = 1$ mgal. The contribution of such a block to the total potential will be $64.6 \times 10^6/\ell(r, \psi_k)$ for ℓ in meters, and T in m^2/sec^2. The factors $f_1(r, \psi) = 5.07\ S(r, \psi)$ and $f_2(r, \psi) = 64.6 \times 10^6/\ell(r, \psi)$ thus describe the sensitivity of the potential to distant zones when the gravity field is represented by gravity anomalies and a surface layer respectively. These two functions are plotted for $r = R$ (earth's surface) in Figure 3.1 and for an altitude of 200 km in Figure 3.2. These graphs show that the sensitivity of the disturbing potential to distant zones is slightly less in absolute value in the case of the surface layer, except for the areas around $\psi = 40°$ and $\psi = 120°$ where the Stokes function passes through the abscissa. This is true both at the earth's surface and at 200 km altitude. In the area within 15° of the sub-satellite point, where both factors are large, the f_2 factor corresponding to the surface layer representation falls off slightly more quickly. Furthermore, this factor remains small, while the Stokes' function grows again in value. Unfortunately, neither function falls quickly to zero and remains negligibly small, so that neither method really has the desired property. Thus, the neglect of far away blocks is never completely justified, since the neglect of these blocks will always bias the solution for the near blocks to some extent. The extent of this bias can be established through numerical experimentation. If the two satellites are close together, then it is reasonable to assume that far away blocks will affect both satellites in approximately the same manner, thus producing almost no net effect on the velocity between them. For this reason, the effect of far away blocks

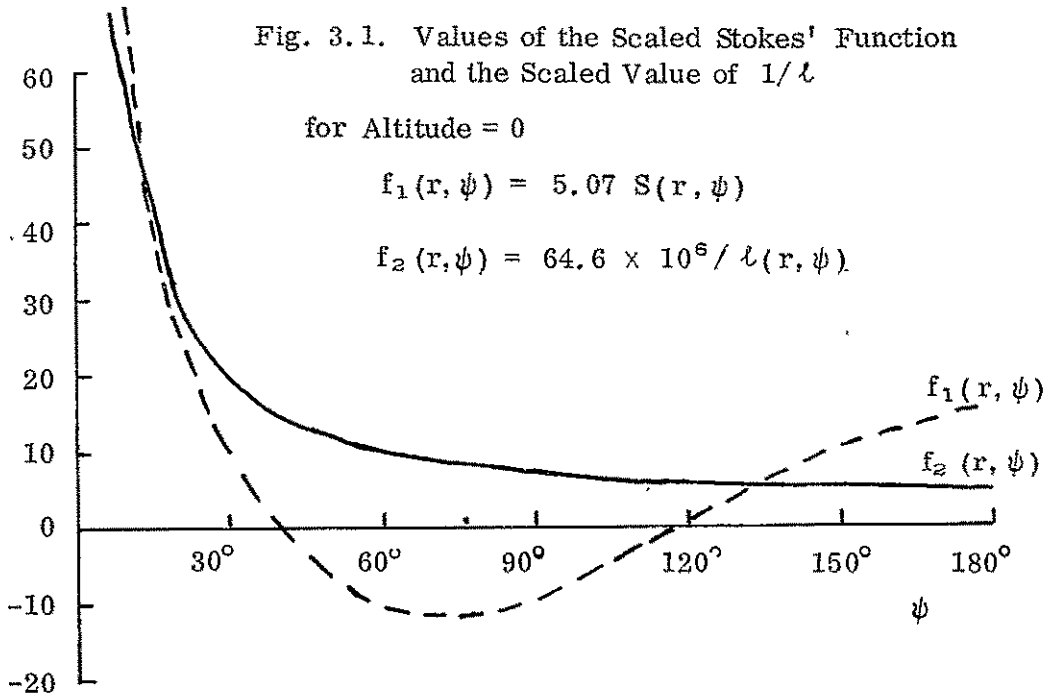

Fig. 3.1. Values of the Scaled Stokes' Function and the Scaled Value of $1/\ell$

for Altitude = 0

$$f_1(r, \psi) = 5.07\ S(r, \psi)$$

$$f_2(r, \psi) = 64.6 \times 10^6 / \ell(r, \psi)$$

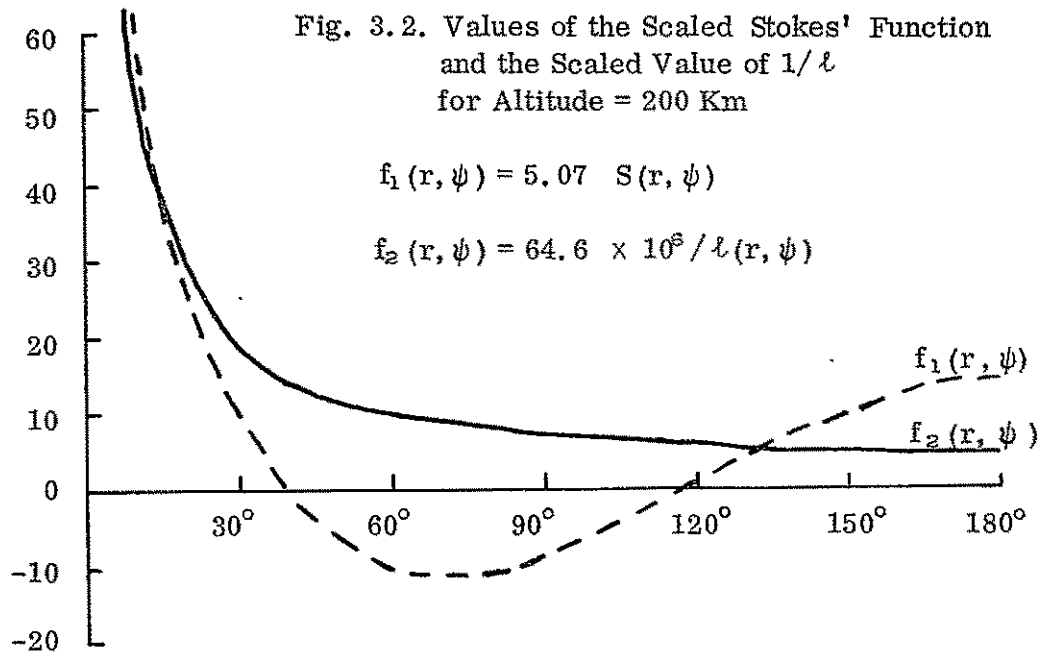

Fig. 3.2. Values of the Scaled Stokes' Function and the Scaled Value of $1/\ell$ for Altitude = 200 Km

$$f_1(r, \psi) = 5.07\ S(r, \psi)$$

$$f_2(r, \psi) = 64.6 \times 10^6 / \ell(r, \psi)$$

was not investigated numerically. In the simulations described in Chapter 5, solutions were made only for the blocks in the vicinity of which the satellite was tracked. This neglect of far away blocks should not invalidate the basic conclusions as to the resolution of the gravity field that can be obtained from satellite to satellite Doppler tracking.

3.3 Transformations Between a Fictitious Surface Layer and Other Representations.

Since the fictitious surface layer is not a common representation of the gravity field, it is desirable to be able to transform the fictitious surface layer into other representations, such as spherical harmonic coefficients or gravity anomalies. The transformation to gravity anomalies is especially important, since any solution for the gravity field made from satellite to satellite range rate data should be compared to independent information obtained from surface gravimetry, where possible. A transformation equation, applicable to a surface layer spread on the regularized geoid to a spherical approximation, is [Heiskanen and Moritz, 1967, p. 303]

$$\Delta g = 2\pi\,\varphi - \frac{3}{2} R \iint_{\sigma} \frac{\varphi}{\ell}\, d\sigma . \qquad (3.5)$$

If the fictitious layer is distributed on the physical surface of the earth S instead of the geoid, then [Heiskanen and Moritz, 1967, p. 302]

$$\Delta g = 2\pi\varphi \cos\beta - \iint_{S} \left(\frac{3}{2 r_p \ell} + \frac{r^2 - r_p^2}{2 r_p \ell^3} \right) \varphi \, dS \qquad (3.6)$$

where $r_p = R + h_p$, $r = R + h$, h_p is the height of the computation point on the surface of the earth, h is height of the integration element dS, and β is the slope of the surface S at the integration element dS. This equation yields the gravity anomaly at the physical surface of the earth, i.e., the difference between actual gravity on the geop passing through a point on the surface and the normal gravity at the corresponding point on

the spherop that has the same potential. This equation also contains a spherical approximation.

It may also be desirable to transform the surface layer representation into a spherical harmonic series representation. This is especially true if a global gravity field described in terms of a fictitious surface layer is to be adjusted to fit the known low order terms of the gravitational potential. By definition, the attraction potential of the surface layer on a particle P at the coordinates $(\varphi'_p, \lambda_p, r_p)$ is

$$T(\varphi'_p, \lambda_p, r_p) = \iint_S \frac{\sigma}{\ell} \, dS \qquad (3.7)$$

where

$$\ell = (r_p^2 + \rho^2 - 2\rho r_p \cos\psi)^{\frac{1}{2}}$$

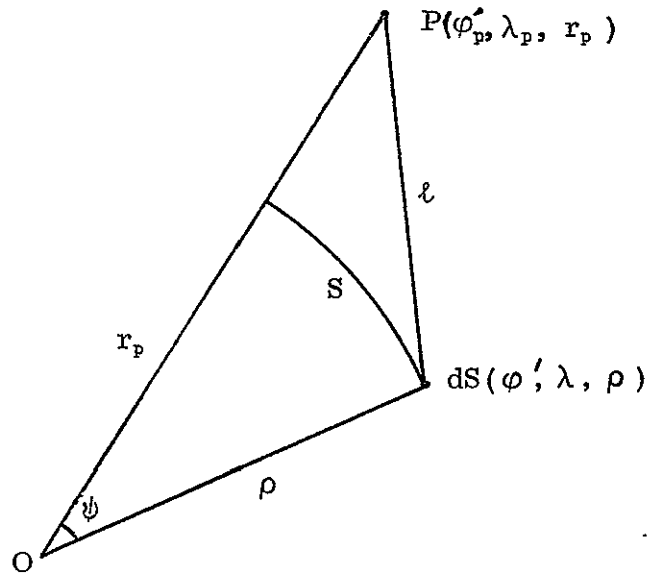

Fig. 3.3 Geometry of the Particle P and the Integration Element dS.

The function $1/\ell$ may be expanded in Legendre polynomials as

$$1/\ell = \frac{1}{r_p} \sum_{n=0}^{\infty} \left(\frac{\rho}{r_p}\right)^n P_n(\cos\psi)$$

where $\cos\psi = \sin\varphi'_p \sin\varphi' + \cos\varphi'_p \cos\varphi' \cos(\lambda - \lambda_p)$. By the decom-

position theorem for spherical harmonics

$$P_n(\cos\psi) = \sum_{m=0}^{n} h_{nm} (\cos m\lambda_p \cos m\lambda + \sin m\lambda_p \sin m\lambda) P_{nm}(\sin\varphi_p') P_{nm}(\sin\varphi')$$

where $$h_{nm} = (2 - \delta_{0m}) \frac{(n-m)!}{(n+m)!} .$$

Substituting this formula into the equation for $1/\ell$, and that equation into (3.7), yields

$$T(\varphi_p', \lambda_p, r_p) = \frac{1}{r_p} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \Big[h_{nm} \iint_S \varphi \left(\frac{\rho}{r_p}\right)^n \cos m\lambda P_{nm}(\sin\varphi') dS \cos m\lambda_p P_{nm}(\sin\varphi_p')$$

$$+ h_{nm} \iint_S \varphi \left(\frac{\rho}{r_p}\right)^n \sin m\lambda P_{nm}(\sin\varphi') dS \sin m\lambda_p P_{nm}(\sin\varphi_p') \Big] .$$

On the other hand, the disturbing potential is conventionally written in terms of spherical harmonics as

$$T(\varphi_p', \lambda_p, r_p) = \frac{GM}{r_p} \sum_{n=0}^{\infty} \left(\frac{a}{r_p}\right)^n \sum_{m=0}^{n} \Big[\Delta C_{nm} \cos m\lambda_p$$

$$+ \Delta S_{nm} \sin m\lambda_p \Big] P_{nm}(\sin\varphi_p') .$$

Here ΔC_{nm}, ΔS_{nm} denote the difference between the coefficients in the total geopotential function W and those in the normal potential U. Identifying the coefficients of the spherical harmonics in the two expressions for T shows that

$$\left.\begin{aligned} \Delta C_{nm} &= \frac{h_{nm}}{GM} \iint_S \varphi \left(\frac{\rho}{a}\right)^n \cos m\lambda P_{nm}(\sin\varphi') dS \\ \Delta S_{nm} &= \frac{h_{nm}}{GM} \iint_S \varphi \left(\frac{\rho}{a}\right)^n \sin m\lambda P_{nm}(\sin\varphi') dS \end{aligned}\right\} \quad (3.8)$$

These expressions are completely rigorous and valid for any surface S.

An approximate inversion of (3.8), by which the density parameter φ may be expressed in terms of the spherical harmonic coefficients, may also be obtained. If the surface S on which the density is distributed is approximated by a sphere of radius R, and we consider $R \approx a$, then

$$\begin{Bmatrix} \Delta C_{nm} \\ \Delta S_{nm} \end{Bmatrix} = h_{nm} \frac{R^2}{GM} \iint_{\sigma} \omega \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} P_{nm}(\sin\varphi') \, d\sigma . \qquad (3.9)$$

Furthermore, the values of φ can be expanded in a series of surface spherical harmonics on this sphere and the function ω can be expressed as a sum of these harmonics, i. e.

$$\varphi(\varphi', \lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} (\varphi_{C_{nm}} \cos m\lambda + \varphi_{S_{nm}} \sin m\lambda) P_{nm}(\sin\varphi') \qquad (3.10)$$

where the coefficients are given by

$$\begin{Bmatrix} \varphi_{C_{nm}} \\ \varphi_{S_{nm}} \end{Bmatrix} = h_{nm} \frac{(2n+1)}{4\pi} \iint_{\sigma} \omega \begin{Bmatrix} \cos m\lambda \\ \sin m\lambda \end{Bmatrix} P_{nm}(\sin\varphi') \, d\sigma \qquad (3.11)$$

Identification of this with equation (3.9) above yields

$$\begin{Bmatrix} \varphi_{C_{nm}} \\ \varphi_{S_{nm}} \end{Bmatrix} = \frac{(2n+1)}{4\pi} \frac{GM}{R^2} \begin{Bmatrix} \Delta C_{nm} \\ \Delta S_{nm} \end{Bmatrix} . \qquad (3.12)$$

Thus

$$\varphi_n = \sum_{m=0}^{n} (\varphi_{C_{nm}} \cos m\lambda + \varphi_{S_{nm}} \sin m\lambda) P_{nm}(\sin\varphi')$$

$$= \frac{2n+1}{4\pi} \frac{GM}{R^2} \sum_{m=0}^{n} (\Delta C_{nm} \cos m\lambda + \Delta S_{nm} \sin m\lambda) P_{nm}(\sin\varphi')$$

$$= \frac{(n + \frac{1}{2})}{2\pi R} T_n . \qquad (3.13)$$

where T_n is the nth degree term in the disturbing potential. Finally

$$\varphi = \sum_{n=2}^{\infty} \varphi_n = \frac{1}{2\pi R} \sum_{n=2}^{\infty} (n + \tfrac{1}{2}) T_n \qquad (3.14)$$

expresses the density of the surface layer in terms of the spherical harmonic coefficients of the disturbing potential. This equation is applicable to the geoid or the ellipsoid to a spherical approximation.

3.4 A Layer Spread on the Surface of the Earth.

The disturbing potential can be produced by a fictitious layer spread on any continuous surface, such as the geoid, an ellipsoid, a sphere, or the physical surface of the earth. Of these, the physical surface of the earth provides the most satisfying interpretation, since the potential and the attraction of such a layer are continuous down to and on the surface [Heiskanen and Moritz, 1967, p. 7]. Thus, such a surface is capable of producing the anomalous attraction measured at the surface of the earth by gravimeters, as well as the anomalous attraction experienced by satellites, without any mathematical complications. The only difficulty is that the geometric shape of the physical surface is not completely known until the shape of the geoid is completely known. However, the major features of the geoid are known. Since the geoid is a rather smooth surface, the model of the geoid derived from a modern gravity field model expressed in terms of spherical harmonic coefficients probably does not differ from the actual geoid by more than a few meters. If such a model is truncated at a low degree and order, such as degree 4, most of the major features of the geoid will still be present [Koch, 1968]. If the SAO 1969 Standard Earth gravity field is truncated at degree 12, and the truncated expression is used to generate a geoid map, the resulting map is practically indistinguishable from that obtained from the full model at the normal 10 meter contour interval. This suggests that such a truncated model could be used as an approximation to the geoid. If the topographic heights are added to this surface, a good approximation to the physical surface of the earth can be described.

This also suggests a somewhat different definition of the normal and disturbing gravity fields. Let

$$W = U' + T' \quad . \tag{3.15}$$

where U' is given by a spherical harmonic series cut off at degree N

$$U' = \frac{GM}{r} \sum_{n=0}^{N} \left(\frac{a}{r}\right)^n \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) \, P_{nm} (\sin\varphi') + \tfrac{1}{2} \omega^2 r^2 \cos^2 \varphi' \qquad (3.16)$$

and the function T' generates the rest of the potential. Since the low degree coefficients are well known, U' may be considered a known function. Quantities computed from the function U' form what is called a "spherop reference system" by Needham [1970]. Since the surface $U' = W_o$ approximates the geoid much better than any ellipsoid [Koch, 1968], the potential contributed by the function T' must be much less than the potential generated by the conventional disturbing potential T. Thus, it would appear that the density of the fictitious layer would be much less if the layer is used to represent T' than if it represents T. Since the main purpose of satellite to satellite tracking is to provide refinement of the gravity field, not to redetermine the low order coefficients, this partitioning of the gravity field was chosen. This partitioning means that the unknown density of the surface layer, when converted into spherical harmonic coefficients by equation (3.8), must not yield any non-zero coefficients of degree N or smaller. This observation imposes $(N + 1)^2$ constraints on the solution for the density of the fictitious surface layer.

The partitioning of the gravity field given by equation (3.15) was observed in the algorithm described in Chapter 4 and in the simulated solutions described in Chapter 5. The new normal potential U' is considered known, while the unknowns describe a surface layer that gives rise to the new disturbing potential T'. Reference orbits are generated using the normal potential U' rather than the conventional normal potential of a level ellipsoid.

3.5 Practical Computations with a Fictitious Surface Layer.

In practical computations, the fictitious layer is represented by blocks in which the density is constant. The density assigned to a block is theoretically the mean density of the surface layer in that block. The interpretation is only slightly different from the case of gravity anomalies, since there is no real surface layer of whose density to take the mean. However, if the density of the fictitious layer were computed from gravity anomalies by equation (3.4), then continuous values could be assigned to the surface layer wherever continuous values of gravity anomalies are measured.

If the density is constant within a block, then all formulae involving integrals of the density can be converted to finite sums. Specifically, the disturbing potential is given by

$$T' = \sum_k \varphi_k \iint \frac{1}{\ell} \, dS_k \qquad (3.17)$$

where the sum is taken over all the blocks and the integral is taken over the area of the kth block S_k. This integral is especially troublesome. Unless the distance of the satellite from the block is very much larger than the dimensions of the block, the quantity $1/\ell$ cannot be treated as constant within the block. In some cases this integral may be evaluated analytically, especially if the portion of the surface contained in each block is sufficiently small that it may be approximated by a plane. It is also possible to evaluate this integral numerically by dividing each block into some number of smaller blocks and assuming the value of $1/\ell$ to be constant within each sub-block. The error of approximation of this approach again depends principally on the ratio of the dimension of the sub-block to the value of ℓ [Needham, 1970]. For the algorithm described in Chapter 4, this integral was always evaluated by dividing the block into four sub-blocks, so that

$$T' = \sum_k \varphi_k \Delta S_k \frac{1}{4} \sum_{i=1}^{4} \frac{1}{\ell_{ik}} \tag{3.18}$$

where ΔS_k is the area of the kth block and ℓ_{ik} is the distance from the computation point (x_p, y_p, z_p) to the center of the ith sub-block of the kth block (x_{ik}, y_{ik}, z_{ik}). The disturbing force at the computation point due to the surface layer is then computed as

$$\nabla T'(x_p, y_p, z_p) = -\sum_k \varphi_k \Delta S_k \frac{1}{4} \sum_{i=1}^{4} \frac{1}{\ell_{ik}^3} \begin{pmatrix} x_p - x_{ik} \\ y_p - y_{ik} \\ z_p - z_{ik} \end{pmatrix} . \tag{3.19}$$

The error of this approximation may be the cause of most of the numerical error detected in the simulated solutions described in Chapter 5.

The specification of the surface S enters these equations through the area ΔS_k and the coordinates (x_{ik}, y_{ik}, z_{ik}). The coordinates of a point on the geoid may be found approximately by computing a value of r_{ik} such that $U'(r_{ik}, \varphi'_{ik}, \lambda_{ik}) = W_o$. The spherical coordinates of a point on the physical surface are then $(r_{ik} + h_{ik}, \varphi'_{ik}, \lambda_{ik})$ where h_{ik} is the topographic height of the point with spherical coordinates $\varphi'_{ik}, \lambda_{ik}$. The topographic heights may be obtained from maps or from lists such as [Kaula, et. al., 1966] or [Strange and Woollard, 1964]. In the simulated solutions described in Chapter 5, the topographic heights were ignored and the fictitious layer was defined to be spread on the surface $U' = W_o$. This is permissible for a study in which no real data is available, but a fully developed data reduction process should consider the topographic heights. The simplification was made in order to reduce the computational effort, and it in no way invalidates the conclusions about the resolution of the gravity field that can be obtained with satellite to satellite tracking.

4. AN ALGORITHM FOR SIMULATING AND ADJUSTING SATELLITE TO SATELLITE RANGE RATE DATA

The algorithm described in this chapter simulates least squares solutions for the parameters describing the gravity field from satellite to satellite range rate data. Since the orbits of the two satellites cannot be considered perfectly known, the orbit elements of each satellite for each orbit also enter the algorithm as unknowns. However, these are considered nuisance parameters of no particular interest. The observed quantities are the range rate between the satellites and the position of each satellite. In practice the position of a satellite is not observed directly; rather, some function of position, such as the range or direction of the satellite from one or more tracking stations on the surface of the earth, is observed. Rather than make any assumption about the mode of tracking from the ground, the algorithm assumes that all three components of the satellite position are observed, and thus generates three observation equations for each simulated position observation of each satellite. In the simulations described in Chapter 5, a low weight was purposely assigned to the satellite position observations to reflect the probability that not all components of position would actually be observed, and that the observations would be made from ground based tracking stations whose geocentric position would not be known with certainty. The weights may be varied, so that a solution based only on the range rate data, with no ground based tracking data whatsoever, can be simulated by assigning zero weight to the position observation equations.

Although many forces physically act on a satellite, only gravitational forces are considered in this section. Thus the force acting on the satellite is represented as the gradient of the gravitational potential of the earth V,

$$F = \nabla V$$

It is assumed that the small luni-solar and non-gravitational forces could

be added to the algorithm as needed should actual satellite to satellite range rate data become available.

As discussed in section 3.4, it is convenient to partition the total potential into normal and disturbing parts V = U + T. Here the symbols V and U denote the same quantities as W and U′ respectively in section 3.4, except that the potential of the centrifugal force is not included. Thus these symbols apply to an inertial, rather than an earth fixed coordinate system. Again the normal potential is represented by a truncated harmonic series,

$$U = \frac{GM}{r} \sum_{n=0}^{N} \left(\frac{a}{r}\right)^n \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) P_{nm}(\sin\psi) \qquad (4.2)$$

where ψ denotes the spherical latitude, λ the longitude, and r the radial coordinate. The second part is the disturbing potential T, which is represented as the potential of a surface layer spread over the surface of the earth. The advantage of this partitioning is that almost all of the potential is put into the normal potential U, and the disturbing potential T is quite small at all points. This means that a reference orbit generated using only the normal potential approximates the actual orbit, and the approximation is sufficiently good that the differences can be represented as linear variations. Since the principal unknowns of the problem are the parameters describing the mean density of the surface layer in some convenient sized blocks, a value of zero may be used as an approximate value of the mean density in all blocks, and the resulting observation equations are sufficiently linear that the final adjusted values of the parameters may be obtained in a single iteration. A considerable saving of labor is achieved, since the surface layer may be completely neglected in all computations leading to the generation of the reference orbit and the partial derivatives.

The algorithm has two functions. First, the "true" orbits of both satellites are integrated numerically for a specified period of time. A disturbing potential represented by a fictitious layer is specified, and the integration takes into account the attraction of both the normal and disturbing components of the field. The range rate between the two satellites is computed

and recorded at desired intervals during the integration, and these quantities then simulate the observed range rates. The positions of both satellites are also recorded at suitable intervals to simulate position observations.

The second function is to form observation equations for each observed quantity. These observation equations are then solved by the least squares algorithm, and the densities of the surface layer specified in the first step are recovered. The elements of each orbit for each satellite may also be recovered, along with the variances and covariances of all recovered quantities.

The algorithm assumes that both satellites are in low orbits and are considerably perturbed by the attraction of the surface layer. If one of the satellites is in a high geostationary orbit, its orbit will not be significantly affected by this attraction. Both the computed orbit for the high satellite and the coefficients of the observation equations will reflect the fact that this effect is insignificant, so that the algorithm is applicable in either case.

4.1 The Integration of the Orbits

The integration of the orbits presents far fewer difficulties than the computation of the observation equations. Let X denote the three components of the position vector and $\dot{X}$ denote the three components of the velocity vector of the satellite in an inertial coordinate system. Then the acceleration is given by

$$\ddot{X} = F = \nabla V = \nabla U + \nabla T \qquad (4.3)$$

The components of ∇T are given in an earth fixed coordinate system by equation (3.19) and may be transformed to the inertial coordinate system by a simple orthogonal linear transformation. The computation of the components of ∇U is described in section (4.23). The use of these equations for direct numerical integration is the Cowell method. It requires a step size that is up to 10 times shorter than that used in the Enke method, which utilizes an elliptic reference orbit. However, the computing time per step

is about 50% less in the Cowell method, and rectification of the reference orbit is not necessary. The Enke method is most suitable for near elliptic orbits, such as orbits far from the earth, and for situations in which a long step size is desirable, such as problems in which the satellite ephemeris is only needed at widely spaced intervals [Conte, 1962]. The Cowell method is chosen for the case of satellite to satellite tracking because at least one of the satellites is quite near to the earth and is considerably perturbed, and because the positions and velocities of the satellites are needed at fairly close intervals so that the range rate may be computed.

The three second order differential equations (4.3) are transformed to a form suitable for numerical integration by setting

$$Y = \begin{pmatrix} X \\ \dot{X} \end{pmatrix},$$

so that

$$\frac{d}{dt} Y = \begin{pmatrix} \dot{X} \\ \ddot{X} \end{pmatrix} = \begin{pmatrix} \dot{X} \\ f(X) \end{pmatrix} = \bar{f}(Y, \varphi_k) . \qquad (4.4)$$

This represents a system of six first order equations. The first three are linear and the last three are non-linear, so that the whole system is non-linear and homogeneous.

These equations may be numerically integrated by any standard method for such systems. The method chosen for these studies uses the modified Hamming fourth order predictor corrector equations [Hamming, 1969]. These equations are incorporated in a detailed integration algorithm described by Ralston in [Ralston and Wilff, 1960, Chapter 8]. The algorithm includes a fourth order Runge Kutta starter. The method is considered stable, and provisions for controlling the local truncation error by changing the step size are included. If the weighted sum of the estimated local truncation errors for all the equations in the system is greater than a preset tolerance, the step size is halved. If the estimated error is still greater than the tolerance, the step size is halved again, up to a total of ten halvings. If the estimated error is less

than one fiftieth of the tolerance, the step size is doubled, except that the step size is never made larger than the initial step size.

If ρ is the distance between the satellites, then $\rho = |X_1 - X_2|$, where the subscripts identify the satellite. The observable quantity is then

$$= \frac{(\dot{X}_1 - \dot{X}_2) \cdot (X_1 - X_2)}{\rho} , \qquad (4.5)$$

which is the projection of the relative velocity vector onto the relative position vector. To compute the range rate at some epoch the positions and velocities of both satellites must be available. To ensure that this will always be so, both orbits are integrated simultaneously. Thus two systems of the form (4.4) are integrated together as a larger system of 12 equations. The initial step size is chosen equal to the desired interval between range rate observations. If the step size is halved by the integration algorithm, then the positions and velocities of the satellites are made available at some epochs for which a simulated observation is not desired, and are ignored. The interval at which the positions of both satellites are recorded as simulated position observations is some integer multiple of the initial step size. Thus no interpolation between integration steps is performed.

4.2 The Observation Equations

Since the orbits are integrated in the Cowell form, the natural choice for the orbit elements is the set of components of the position and velocity vectors at the intial epoch, $X_o \equiv X(t_o)$ and $\dot{X}_o \equiv \dot{X}(t_o)$. The observed range rate between the satellites is written functionally in terms of the unknown parameters as

$$\dot{\rho}(t) = \dot{\rho}(X_{o1}, \dot{X}_{o1}, X_{o2}, \dot{X}_{o2}, \varphi_K ; t) \qquad (4.6)$$

The coefficients of the observation equations are the partial derivatives of $\dot{\rho}$ with respect to each of the unknowns, evaluated at time t with approximate values of the unknowns. The range rate $\dot{\rho}(t)$ depends explicitly only on the

positions and velocities of both satellites at time t. The constant term in the observation equation requires an approximate value $\dot{\rho}_N$ of the range rate, computed from approximate values of the unknowns.

For either orbit, let $Y_{NO} \equiv (X_{NO}, \dot{X}_{NO})$ denote a set of initial epoch elements that approximate those of the unknown real orbit, and let ω_{NK} denote a set of values for the mean densities that approximate the unknown real values. By the discussion at the beginning of this chapter, all the ω_{NK} are taken to be zero. Thus the force function in the gravity field characterized by the φ_{NK} is simply the gradient of the normal potential U; the disturbing potential T is identically zero. Let an orbit be integrated in this normal field with initial epoch conditions Y_{NO}. The position and velocity components of this nominal orbit are denoted by $Y_N \equiv (X_N, \dot{X}_N)$. The range rate $\dot{\rho}(t)$ depends explicitly only on the positions and velocities of both satellites at time t, so that the computed observable $\dot{\rho}_N(t)$ is obtained by evaluating (4.5) with the elements $Y_N(t)$ for each satellite.

The position and velocity elements of the first satellite at time t depend in turn on the epoch elements for that satellite $(X_{o1}, \dot{X}_{o1})$ as well as on the mean densities in blocks of the fictitious surface layer φ_K. Thus

$$\frac{\partial \dot{\rho}(t)}{\partial X_{o1}} = \frac{\partial \dot{\rho}(t)}{\partial X_1(t)} \frac{\partial X_1(t)}{\partial X_{o1}} + \frac{\partial \dot{\rho}(t)}{\partial \dot{X}_1(t)} \frac{\partial \dot{X}_1(t)}{\partial X_{o1}} , \qquad (4.7)$$

with similar equations for the other sets of initial epoch unknowns $\dot{X}_{o1}$, X_{o2}, and $\dot{X}_{o2}$. Also,

$$\frac{\partial \dot{\rho}(t)}{\partial \varphi_K} = \frac{\partial \dot{\rho}(t)}{\partial X_1(t)} \frac{\partial X_1(t)}{\partial \varphi_K} + \frac{\partial \dot{\rho}(t)}{\partial \dot{X}_1(t)} \frac{\partial \dot{X}_1(t)}{\partial \varphi_K} + \frac{\partial \dot{\rho}(t)}{\partial X_2(t)} \frac{\partial X_2(t)}{\partial \varphi_K} + \frac{\partial \dot{\rho}(t)}{\partial \dot{X}_2(t)} \frac{\partial \dot{X}_2(t)}{\partial \varphi_K} . \qquad (4.8)$$

The sets of partial derivatives $\frac{\partial \dot{\rho}}{\partial X_1}$, $\frac{\partial \dot{\rho}}{\partial \dot{X}_1}$, etc., where the explicit reference to time t is dropped, are obtained by differentiating equation (4.5),

$$\frac{\partial \dot{\rho}}{\partial X_1} = \frac{\dot{X}_1 - \dot{X}_2}{\rho} - \frac{\dot{\rho}(X_1 - X_2)}{\rho^2}$$

$$\frac{\partial \dot{\rho}}{\partial \dot{X}_1} = \frac{X_1 - X_2}{\rho}$$

$$\frac{\partial \dot{\rho}}{\partial X_2} = -\frac{\partial \dot{\rho}}{\partial X_1}$$

$$\frac{\partial \dot{\rho}}{\partial \dot{X}_2} = -\frac{\partial \dot{\rho}}{\partial \dot{X}_1} \qquad (4.9)$$

The matrix of partial derivatives

$$\Phi = \begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix} = \begin{pmatrix} \varphi_1 & \varphi_2 \\ \varphi_3 & \varphi_4 \end{pmatrix} = \begin{pmatrix} \frac{\partial X}{\partial X_o} & \frac{\partial X}{\partial \dot{X}_o} \\ \frac{\partial \dot{X}}{\partial X_o} & \frac{\partial \dot{X}}{\partial \dot{X}_o} \end{pmatrix} \qquad (4.10)$$

is called the state transition matrix. It describes the transition of a differential variation of the initial epoch conditions from time t_o to time t, i.e.,

$$\delta Y = \delta \begin{pmatrix} X \\ \dot{X} \end{pmatrix} = \Phi \, \delta \begin{pmatrix} X_o \\ \dot{X}_o \end{pmatrix} = \Phi \, \delta Y_o .$$

Such a matrix exists for each of the satellites. Their elements are needed in the equations of the form (4.7). The matrix of partial derivatives

$$W = \begin{pmatrix} \frac{\partial X}{\partial \varphi_K} \\ \frac{\partial \dot{X}}{\partial \varphi_K} \end{pmatrix} = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \qquad (4.11)$$

is called the parameter sensitivity matrix. It describes the effect of a differential variation in the parameters φ_K on the orbit,

$$\delta Y = W \, \delta \varphi_K .$$

Such a matrix also exists for each of the orbits, and their elements are needed to evaluate equation (4.8). The net effect of variations in all the parameters describing a single orbit is given by

$$\delta Y = \Phi \, \delta Y_o + W \, \delta \varphi_K . \qquad (4.12)$$

These equations may also be viewed as a linearization of the equation

$Y = Y(Y_o, \varphi_K)$ by a Taylor series expansion truncated at the first degree,

$$Y = Y_N + \frac{\partial Y}{\partial Y_o}(Y_o - Y_{NO}) + \frac{\partial Y}{\partial \varphi_K}(\varphi_K - \varphi_{NK}).$$

This identifies the variations δY and δY_o in (4.12) as the differences between the real elements of the orbit and those of the nominal orbit. It also makes it clear that the partial derivatives are to be evaluated along the nominal orbit.

Variations may also be taken of the equations of motion (4.4),

$$\frac{d}{dt}(\delta Y) = \bar{\bar{G}}\delta Y + \bar{\bar{H}}\delta\varphi_K \qquad (4.13)$$

where

$$\bar{\bar{G}} = \frac{\partial \bar{f}}{\partial Y}(Y)$$

$$\bar{\bar{H}} = \frac{\partial \bar{f}}{\partial \varphi_k}(Y)$$

are matrices of partial derivatives which are also to be evaluated along the nominal orbit. Substituting (4.12) into the variational equations (4.13) yields

$$\frac{d}{dt}\Phi\,\delta Y_o + \frac{d}{dt}W\delta\varphi_K = \bar{\bar{G}}\,\Phi\,\delta Y_o + \bar{\bar{G}}\,W\delta\varphi_K + \bar{\bar{H}}\,\delta\varphi_K.$$

This must hold for all variations δY_o and $\delta\varphi_K$, which implies the following two matrix systems of differential equations:

$$\frac{d}{dt}\Phi = \bar{\bar{G}}\,\Phi, \qquad (4.14)$$

$$\frac{d}{dt}W = \bar{\bar{G}}\,W + \bar{\bar{H}}. \qquad (4.15)$$

The elements of the matrices $\bar{\bar{G}}$ and $\bar{\bar{H}}$ depend explicitly only on the φ_{NK} and on the elements of the nominal orbit, so that they are implicitly functions only of the independent variable, the time. Thus the matrix system (4.14) is linear and homogeneous and the system (4.15) is linear and non-homogeneous. Since at the initial epoch t_o we must have $\delta Y(t_o) = \delta Y_o$,

the initial conditions for (4.14) are $\Phi(t_o) = I$, the identity matrix, and the initial conditions for (4.15) are $W(t_o) = 0$. The total set of equations to be integrated is now

$$\frac{d}{dt}\begin{pmatrix} Y_N \\ \Phi \\ W \end{pmatrix} = \begin{pmatrix} \bar{f}(Y_{NO}, \varphi_{NK}) \\ \bar{G}\ \Phi \\ \bar{G}\ W + \bar{H} \end{pmatrix} \qquad (4.16)$$

with initial conditions

$$\begin{pmatrix} Y_o \\ I \\ 0 \end{pmatrix} .$$

The set Y_N contains six elements, Φ contains 36 elements, and the number of elements in W is six times the number of parameters φ_K that describe the disturbing potential. It is interesting to note that all of the equations are linear except the second group of three in the set Y_N; however, this is sufficient to make the whole system non-linear.

If, in the case of satellite to satellite range rate tracking, both satellites are in low orbits and are significantly perturbed by the disturbing force, it might be necessary to integrate a system of the form (4.16) for each satellite. The simultaneous numerical integration of two such systems is certainly possible, and this approach was numerically tested during the design of the algorithm. However, the requirement for computer core memory becomes a limiting factor if the number of parameters φ_K is large, and it is worthwhile investigating other methods of obtaining the Φ and W matrices.

4.21 The State Transition Matrix. The transition matrix Φ is sometimes referred to as the matrizant of the orbit, or of the matrix system of linear differential equations (4.14). Strictly speaking, the transition matrix is the matrizant of the matrix of coefficients $\bar{G}$. This terminology corresponds to an integral representation of the state transition matrix which

is interesting but not especially useful [Schwarz, 1967].

Another alternative is to consider the fact that the transition matrix consists of partial derivatives, and thus can be obtained by numerical differentiation. In addition to the nominal orbit, six other orbits are integrated, each of the six corresponding to a small variation in one of the initial conditions. The differences between the elements of these orbits and those of the nominal orbits are variations of the elements. The ratios of these variations at any time epoch to the variations in initial conditions approximate the partial derivatives (4.10). Since each of the six variational orbits involves six coordinates, it is necessary to integrate 36 equations in addition to the nominal orbit, which is the same number that would be involved were the differential equation (4.14) for Φ integrated directly. Furthermore, considerable numerical experience is necessary to determine the proper variation in initial conditions to use. The numerical differentiation approach may be used also in the case of the parameter sensitivity matrix W, and the same remarks on its use apply.

A third alternative is to use a state transition matrix computed for an orbit simpler than the nominal orbit. This usually means that the desired transition matrix is approximated by the transition matrix that belongs to the Keplerian elliptic orbit having the same initial conditions as the nominal orbit. The transition matrix for an elliptic orbit can be written explicitly as a function of time in a form involving only elementary functions, so that a considerable saving of effort may be achieved. On the other hand, experience is required to judge the conditions under which the approximation is satisfactory. For considerably perturbed orbits near the earth, the approximation will only be satisfactory for a small fraction of a revolution.

The direct numerical integration of equation (4.14) was selected as an appropriate method of generating Φ for this algorithm. This integration is now examined in greater detail.

As mentioned previously, equation (4.14) is formally a linear differential

equation, since the matrix $\bar{G}$ depends only on the independent variable (time), and not on the dependent variable Φ. However, the dependence of $\bar{G}$ on t is not explicit, but implicit through the state variables Y_N, which are in turn functions of time. Since the equations of motion are integrated numerically the matrix $\bar{G}$ cannot be written as a continuous function of the time. Instead, discreet values of $\bar{G}$ can be computed only at those epochs for which values of the state variables Y_N are computed. Because of this property, the step sizes used in integrating $\dot{\Phi} = G\,\Phi$ must be identical to (or integer multiples of) the step sizes used in integrating the nominal trajectory. Otherwise, it would be necessary to interpolate in the ephemeris of the trajectory for each epoch for which a value of G is desired. In order to avoid the need for such interpolation, it is convenient to integrate the transition matrix along with the trajectory, letting the accuracy required in the trajectory determine the step size. This means that the equations for the trajectory and the transition matrix are integrated together as one large system of non-linear differential equations, and the advantages that might be taken of the fact that the equation for Φ is linear are neglected. On the other hand, the same block of coding is used for integrating both the trajectory and the transition matrix, resulting in a saving of machine time and space. Another possible objection is that there is no control on the local truncation error in the integration of the transition matrix; however, the variational equations are generally better behaved than the trajectory equations, and usually need not satisfy as stringent accuracy requirements [Riley, et al., 1967]. Therefore, the step size selected for the integration of the trajectory is almost always also appropriate for the integration of the transition matrix. These remarks also apply to the differential equation for the parameter sensitivity matrix W.

The 6 x 6 matrix $\bar{G}$ may be expressed in terms of 3 x 3 submatrices by considering the definitions of Y and $\bar{f}$ (Y) in equation (4.4). Thus

$$\bar{\bar{G}} = \frac{\partial \bar{f}(Y)}{\partial Y} = \frac{\partial(\dot{X}, \ddot{X})}{\partial(X, \dot{X})}$$

Since only gravitational forces are being considered, the acceleration depends only on the position elements X and is independent of the velocity elements $\dot{X}$. Thus

$$\bar{\bar{G}} \quad \begin{pmatrix} \frac{\partial \dot{X}}{\partial X} & \frac{\partial \dot{X}}{\partial \dot{X}} \\ \frac{\partial \ddot{X}}{\partial X} & \frac{\partial \ddot{X}}{\partial \dot{X}} \end{pmatrix} = \begin{pmatrix} 0 & I \\ G & 0 \end{pmatrix} \qquad (4.17)$$

where

$$G = \frac{\partial f(X)}{\partial X} \qquad (4.18)$$

The detailed computation of the G matrix as a function of the state variables Y_N is shown in section 4.3.

4.22 The Parameter Sensitivity Matrix. As noted earlier, it is possibl to integrate all three groups of differential equations (trajectory, transition matrix, and parameter sensitivity matrix) together as a single non-linear system of equations. On the other hand, it is also possible to express W as a definite integral, and evaluate it for any given time by an appropriate method for numerical integration of definite integrals. The definite integral expression for W is arrived at by expressing the solution as the sum of complementary and particular integrals, which is the standard method for solving non-homogeneous linear differential equations. The homogeneous system of equations corresponding to the non-homogeneous system (4.15) is

$$\frac{d}{dt} W = \bar{\bar{G}} W.$$

This is the same as the differential equation for Φ, so that the complementary solution is

$$W(t) = \Phi(t)\, W(t_o).$$

The matrix $W(t_o)$ bears the burden of the constants of integration. In the

case of the parameter sensitivity matrix, these are $W(t_o) = 0$. The Green's matrix of the system is $\Phi(t)\ \Phi^{-1}(\tau)$, and the particular integral is given by

$$\int_{t_o}^{t} \Phi(t)\ \Phi^{-1}(\tau)\ \bar{H}(\tau)\ d\tau.$$

Adding the complementary and particular integrals and using the initial conditions $W(t_o) = 0$, the complete solution is

$$W(t) = \Phi(t) \int_{t_o}^{t} \Phi^{-1}(\tau)\ \bar{H}(\tau)\ d\tau. \qquad (4.19)$$

The property that makes the use of this expression desirable is that Φ^{-1} can be expressed easily in terms of Φ, considerably simplifying the integrand. The development below follows that of Danby [1962]. An entirely different development leading to the same results is given in [Warner and Nead, 1965].

Using the partitioning indicated in equation (4.10), write

$$\Phi = \begin{pmatrix} \varphi_1 & \varphi_2 \\ \varphi_3 & \varphi_4 \end{pmatrix} \quad \text{and} \quad \Phi^{-1} = \Gamma = \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_4 \end{pmatrix}.$$

With

$$\bar{G} = \begin{pmatrix} 0 & I \\ G & 0 \end{pmatrix}$$

the following differential equations may be extracted from equation (4.14):

$$\begin{aligned} \dot{\varphi}_1 &= \varphi_3, & \varphi_1(t_o) &= I, \\ \dot{\varphi}_2 &= \varphi_4, & \varphi_2(t_o) &= 0, \\ \dot{\varphi}_3 &= G\varphi_1, & \varphi_3(t_o) &= 0, \\ \dot{\varphi}_4 &= G\varphi_2, & \varphi_4(t_o) &= I. \end{aligned}$$

Differentiating the first two equations with respect to time and substituting into the second two equations yields the second order differential equations

$$\begin{aligned} \ddot{\varphi}_1 &= G\varphi_1, & \varphi_1(t_o) &= I, & \dot{\varphi}_1(t_o) &= 0; \\ \ddot{\varphi}_2 &= G\varphi_2, & \varphi_2(t_o) &= 0, & \dot{\varphi}_2(t_o) &= I. \end{aligned}$$

Also, differentiating $\Gamma\ \Phi = \Phi^{-1}\Phi = I$ with respect to time

$$\dot{\Gamma}\,\Phi + \Gamma\,\dot{\Phi} = 0.$$

Substituting $\dot{\Phi} = \bar{G}\,\Phi$ and solving for $\dot{\Gamma}$,

$$\dot{\Gamma}\,\Phi + \Gamma\,\bar{G}\,\Phi = 0,$$

$$\dot{\Gamma} = -\Gamma\bar{G}.$$

Extracting submatrices from this equation and using the initial conditions $\Gamma(t_o) = [\Phi(t_o)]^{-1} = I$, the following system is obtained:

$$\begin{aligned}
\dot{\gamma}_1 &= -\gamma_2 G, & \gamma_1(t_o) &= I,\\
\dot{\gamma}_2 &= -\gamma_1\ , & \gamma_2(t_o) &= 0,\\
\dot{\gamma}_3 &= -\gamma_4 G, & \gamma_3(t_o) &= 0,\\
\dot{\gamma}_4 &= -\gamma_3\ , & \gamma_4(t_o) &= I.
\end{aligned}$$

Since the force acting on the satellite is derivable from a scalar potential function V, $\ddot{x}_i = \dfrac{\partial V}{\partial x_i}$, and $(G)_{ij} = \dfrac{\partial \ddot{x}_i}{\partial x_j} = \dfrac{\partial^2 V}{\partial x_i\ \partial x_j}$. Thus $(G)_{ij} = (G)_{ji}$ and $G = G^T$; i.e., the G matrix is symmetric. Using this property and transposing the first and third equations yields

$$\dot{\gamma}_1^T = -G\gamma_2^T,$$

$$\dot{\gamma}_3^T = -G\gamma_4^T.$$

Differentiation of the second and fourth equations gives

$$\dot{\gamma}_1 = -\ddot{\gamma}_2,$$

$$\dot{\gamma}_3 = -\ddot{\gamma}_4,$$

and after substitution,

$$-\ddot{\gamma}_2^T = -G\gamma_2^T,$$

$$-\ddot{\gamma}_4^T = -G\gamma_4^T.$$

Changing signs and indicating initial conditions,

$$(-\ddot{\gamma}_2^T) = G(-\gamma_2^T), \qquad (-\gamma_2^T(t_o)) = 0, \qquad (-\dot{\gamma}_2^T(t_o)) = \gamma_1(t_o) = I;$$

$$\dot{\gamma}_4^T = G\gamma_4^T, \qquad \gamma_4(t_o) = I, \qquad \dot{\gamma}_4(t_o) = -\gamma_3(t_o) = 0.$$

Since $(-\gamma_2^T)$ satisfies the same differential equation as φ_2, and both have the same initial conditions, the two matrices must be identical; i.e.,

$$-\gamma_2^T = \varphi_2 ,$$
$$\text{or} \quad \gamma_2 = -\varphi_2^T .$$

Similarly, γ_4^T must be identical to φ_1, or $\gamma_4 = \varphi_1^T$.

Then $$\gamma_1 = -\dot{\gamma}_2 = \dot{\varphi}_2^T = \varphi_4^T ,$$

and $$\gamma_3 = -\dot{\gamma}_4 = -\dot{\varphi}_1^T = -\varphi_3^T .$$

Finally, collecting the results

$$\Gamma = \Phi^{-1} = \begin{pmatrix} \varphi_4^T & -\varphi_2^T \\ -\varphi_3^T & \varphi_1^T \end{pmatrix} \tag{4.20}$$

A further simplification of the equation for W may be made by noting that

$$\bar{H} = \frac{\partial \bar{f}(Y)}{\partial \varphi_K} = \frac{\partial(\dot{X}, \ddot{X})}{\partial \varphi_K} = \begin{pmatrix} 0 \\ \frac{\partial \ddot{X}}{\partial \varphi_K} \end{pmatrix} = \begin{pmatrix} 0 \\ H \end{pmatrix} \tag{4.21}$$

where $$H = \frac{\partial f(X)}{\partial \varphi_K} . \tag{4.22}$$

Then $$\Phi^{-1}\bar{H} = \begin{pmatrix} -\varphi_2^T H \\ \varphi_1^T H \end{pmatrix} . \tag{4.23}$$

The computation of the elements of the H matrix is discussed in section 4.3.

The final expression for W in (4.19) is

$$W(t) = \begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix} = \begin{pmatrix} \varphi_1(t) & \varphi_2(t) \\ \varphi_3(t) & \varphi_4(t) \end{pmatrix} \int_{t_o}^{t} \begin{pmatrix} -\varphi_2^T(\tau) H(\tau) \\ \varphi_1^T(\tau) H(\tau) \end{pmatrix} d\tau . \tag{4.24}$$

The evaluation of the integral in this equation presents some interesting problems.

Methods for the evaluation of definite integrals require that the integrand

be evaluated at certain epochs over the total interval of integration. However, the elements of the H matrix are functions of the trajectory, and the epochs at which values of Φ are computed also depend on the trajectory. Thus the integrand may be computed without interpolation only at those epochs for which a point on the trajectory is computed. This is in turn controlled both by the selection of the initial step size and by the action of integration module, which may adjust the step size to control the local truncation error in the computation of the trajectory.

It is desirable to evaluate the definite integral along with the numerical integration process. Let $\Psi(t)$ denote the integral in (4.24) and let $\Psi'(\tau)$ denote its integrand. Assume that the integral $\Psi(t_{i-1})$ has been computed and stored in the computer. Also assume that $\Psi'(t_{i-1})$ has been stored and is available. Now let the numerical integration proceed another step to t_i, so that $Y_N(t_i)$ and $\Phi(t_i)$ are available. The desired value of Ψ is then

$$\Psi(t_i) = \Psi(t_{i-1}) + \int_{t_{i-1}}^{t_i} \Psi'(\tau)\, d\tau . \qquad (4.25)$$

If the parameter sensitivity matrix for time t_i is desired, it is easily evaluated as $W(t_i) = \Phi(t_i)\, \Psi(t_i)$; otherwise this computation is ignored.

Equation (4.25) is most easily evaluated by the trapezoid rule: Denoting $\Psi(t_i)$ by Ψ_i

$$\Psi_i = \Psi_{i-1} + \tfrac{1}{2}[\Psi'_{i-1} + \Psi'_i](t_i - t_{i-1}) . \qquad (4.26)$$

The use of this equation in the algorithm was numerically tested. For the step size being used to integrate the orbit, it proved to be insufficiently accurate, leading to an accuracy of only one or two significant digits in the observation equation coefficients.

Simpson's rule provides a significant increase in accuracy over the trapezoid rule. Suppose that the integration has proceeded an even number of steps to time t_i, and that Ψ_{i-2}, Ψ'_{i-2}, and Ψ'_{i-1} have been computed and saved. Then Ψ'_i may be computed from $Y_N(t_i)$ and $\Phi(t_i)$.

Further assume that the last two steps have been of the same size, so that $h = t_i - t_{i-1} = t_{i-1} - t_{i-2}$. Then Simpson's rule states

$$\Psi_i = \Psi_{i-2} + \frac{h}{3}\left[a_1 \Psi'_{i-2} + a_2 \Psi'_{i-1} + a_3 \Psi'_i\right] \quad (4.27)$$

where $a_1 = 1$, $a_2 = 4$, and $a_3 = 1$. This can easily be worked into an algorithm that assures that the current value of Ψ is always available: If the number of steps is odd, Simpson's rule cannot be applied and $\Psi(t)$ must be computed using the trapezoid rule. After the next step, the number of steps is again even. The integral computed by the trapezoid rule is subtracted and the integral over the last two steps is computed by Simpson's rule and added. Since the value of i has been incremented by one, the value of Ψ computed by the trapezoid rule is now denoted Ψ_{i-1}. If only the most recent value of Ψ is saved,

$$\Psi_i = \Psi_{i-1} + \frac{h}{3}\left[(a_1 - \frac{3}{2})\ \Psi'_{i-2} + (a_2 - \frac{3}{2})\ \Psi'_{i-1} + a_3 \Psi'_i\right]. \quad (4.28)$$

The only problem with this algorithm occurs if the step size is changed. Let $h = t_{i-1} - t_{i-2}$ and $g = t_i - t_{i-1}$. If the step sized is halved $h = 2g$, and if it is doubled $2h = g$. Simpson's rule is not strictly applicable, since it requires that the two intervals be equal. However, it is possible to perform an integration having some of the properties of Simpson's rule.

Simpson's rule may be derived by computing the area under the parabola that passes through three points whose abscissae are equally spaced. Thus the rule is exact if the function being integrated is indeed a polynomial of degree two or less. The surprising thing about this rule, and the property that accounts for its popularity, is that the coefficient of the third derivative vanishes in the error term, so that the error depends on the fourth derivative [Conte, 1965]. Thus Simpson's rule is exact even for a cubic polynomial, even though only three points are used.

A modified Simpson's rule may be obtained by computing the area under the parabola that passes through three points whose abscissae are not equally spaced. If this is done, the coefficients in equation (4.27) are given by

$a_1 = \frac{9}{8}$, $a_2 = \frac{27}{8}$, $a_3 = 0$ in the case h = 2g; and by $a_1 = 0$, $a_2 = \frac{27}{4}$, $a_3 = \frac{9}{4}$, in the case 2h = g. These modified rules do not have the advantageous error property of the regular Simpson's rule, so that some accuracy may be lost when the step size is changed. On the other hand, experience indicates that the step size is only seldom changed, so that the regular Simpson's rule is used for almost all steps. These modified rules are still more accurate than two steps using the trapezoid rule.

This part of the algorithm is completed by allowing for a change in step size. The algorithm detects a change in step size and selects the appropriate value of the coefficients a_1, a_2, a_3 for use in equation (4.28). Tests performed with the algorithm in this form indicated that satisfactory values could be obtained for the elements of W with the step size selected for the orbit, although these values were not as accurate as those that could be obtained by direct integration of the differential equation (4.15) for the parameter sensitivity matrix. The coefficients obtained by the two different methods generally agreed to about five significant digits. The advantages of using the algorithm based on the definite integral expression for W are that it is more efficient in terms of both computer time and computer memory space.

4.23 The F, G, and H Matrices. At each step of the integration, it is necessary to compute the force per unit mass $F = \nabla V = \ddot{X} = f(X, \varphi_k)$, as well as the matrices

$$G = \frac{\partial f(X)}{\partial X} \quad \text{and} \quad H = \frac{\partial f(X)}{\partial \varphi_k}$$

For the integration of the nominal orbit, the force vector is the gradient of the normal potential U. This function is expressed in terms of the earth fixed spherical coordinates (λ, Ψ, r), while the components of the force vector in inertial cartesian coordinates are needed. The two coordinate systems are related by

$$\left.\begin{aligned} x &= r \cos \Psi \cos(\lambda+\theta) \\ y &= r \cos \Psi \sin(\lambda+\theta) \\ z &= r \sin \Psi \end{aligned}\right\} \qquad (4.29)$$

where θ denotes the Greenwich sidereal time. Applying the chain rule for partial differentiation, F is obtained by

$$F = \nabla U = \frac{\partial U}{\partial(\lambda, \psi, r)} \quad \frac{\partial(\lambda, \psi, r)}{\partial(x, y, z)} .$$

A greater compactness of notation is achieved by using component notation. Thus, if the symbol x denotes an element of the set (x, y, z) and the symbol y denotes an element of the set (λ, ψ, r), the force may be written

$$F_i = \nabla_i U = \frac{\partial U}{\partial y_j} \frac{\partial y_j}{\partial x_i} \tag{4.30}$$

where the summation convention is used to imply summation over subscripts which appear twice. The components of the vector $\frac{\partial U}{\partial y_j}$ are

$$\left.\begin{aligned}
\frac{\partial U}{\partial \lambda} &= \frac{GM}{r} \sum_{n=0}^{N} \left(\frac{a}{r}\right)^n \sum_{m=0}^{n} m \left[-C_{nm} \sin m\lambda + S_{nm} \cos m\lambda\right] P_{nm}(\sin \psi) \\
\frac{\partial U}{\partial \psi} &= \frac{GM}{r} \sum_{n=0}^{N} \left(\frac{a}{r}\right)^n \sum_{m=0}^{n} \left[C_{nm} \cos m\lambda + S_{nm} \sin m\lambda\right] \frac{d P_{nm}(\sin \psi)}{d \psi} \\
\frac{\partial U}{\partial r} &= \frac{GM}{r^2} \sum_{n=0}^{N} (n+1) \left(\frac{a}{r}\right)^n \sum_{m=0}^{n} \left[C_{nm} \cos m\lambda + S_{nm} \sin m\lambda\right] P_{nm}(\sin \psi)
\end{aligned}\right\} \tag{4.31}$$

The components of $\frac{\partial y_j}{\partial x_i}$ are easily obtained from geometrical considerations or by differentiation of (4.29). Arranged in matrix form they are

$$\begin{pmatrix}
-\frac{\sin(\lambda+\theta)}{r \cos \psi} & \frac{\cos(\lambda+\theta)}{r \cos \psi} & 0 \\
-\frac{\sin \psi \cos(\lambda+\theta)}{r} & -\frac{\sin \psi \sin(\lambda+\theta)}{r} & \frac{\cos \psi}{r} \\
\cos \psi \cos(\lambda+\theta) & \cos \psi \sin(\lambda+\theta) & \sin \psi
\end{pmatrix} . \tag{4.32}$$

In the same notation, the components of the G matrix are

$$G_{ij} = \frac{\partial F_i}{\partial x_j} = \frac{\partial^2 U}{\partial x_i \partial x_j} = \frac{\partial^2 U}{\partial y_k \partial y_p} \frac{\partial y_k}{\partial x_i} \frac{\partial y_p}{\partial x_j} + \frac{\partial U}{\partial y_k} \frac{\partial^2 y_k}{\partial x_i \partial x_j} .$$

The most difficult term to evaluate in this expression is $\frac{\partial^2 y_K}{\partial x_i \partial x_j}$, since this is a triply subscripted quantity, containing 27 elements, only eight of which are zero. However, a better expression for the components of G may be obtained by using the tensor calculus. The components of the force vector are the covariant components of a tensor, and the elements of the G matrix are the components of the covariant derivative of the force. Thus these elements must transform from spherical to Cartesian coordinates as a doubly covariant tensor. In spherical coordinates, the components of the covariant derivative of the force are given by

$$\frac{\partial^2 U}{\partial y_p \partial y_k} - \left\{ \begin{matrix} s \\ p\ k \end{matrix} \right\} \frac{\partial U}{\partial y_s}$$

where $\left\{ \begin{matrix} s \\ p\ k \end{matrix} \right\}$ is the Christoffel symbol of the second kind. Although the Christoffel symbol is also a triply indexed quantity containing 27 elements, only nine of these elements are non-vanishing, and even the non-vanishing elements are simple expressions. This tensor transforms to Cartesian coordinates by the transformation law

$$\left(\frac{\partial^2 U}{\partial x_i \partial x_j} - \left\{ \begin{matrix} \overline{q} \\ i\ j \end{matrix} \right\} \frac{\partial U}{\partial x_q} \right) = \left(\frac{\partial^2 U}{\partial y_p \partial y_k} - \left\{ \begin{matrix} s \\ p\ k \end{matrix} \right\} \frac{\partial U}{\partial y_s} \right) \frac{\partial y_p}{\partial x_i} \frac{\partial y_k}{\partial x_j}$$

where $\left\{ \begin{matrix} \overline{q} \\ i\ j \end{matrix} \right\}$ denotes the Christoffel symbol in Cartesian coordinates. Since the Christoffel symbols in Cartesian coordinates all vanish, this reduces to

$$G_{ij} = \frac{\partial^2 U}{\partial x_i \partial x_j} = \left(\frac{\partial^2 U}{\partial y_p \partial y_k} - \left\{ \begin{matrix} s \\ p\ k \end{matrix} \right\} \frac{\partial U}{\partial y_s} \right) \frac{\partial y_p}{\partial x_i} \frac{\partial y_k}{\partial x_j} . \qquad (4.33)$$

In the spherical coordinates (λ, ψ, r), the non-zero Christoffel symbols are

$$\begin{Bmatrix} 1 \\ 1\ 2 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 2\ 1 \end{Bmatrix} = -\tan\psi,$$

$$\begin{Bmatrix} 1 \\ 1\ 3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 3\ 1 \end{Bmatrix} = \frac{1}{r},$$

$$\begin{Bmatrix} 2 \\ 1\ 1 \end{Bmatrix} = \cos\psi \sin\psi,$$

$$\begin{Bmatrix} 2 \\ 2\ 3 \end{Bmatrix} = \begin{Bmatrix} 2 \\ 3\ 2 \end{Bmatrix} = \frac{1}{r},$$

$$\begin{Bmatrix} 3 \\ 1\ 1 \end{Bmatrix} = -r\cos^2\psi,$$

$$\begin{Bmatrix} 3 \\ 2\ 2 \end{Bmatrix} = -r.$$

The elements of $\frac{\partial U}{\partial y_k \partial y_p}$ are given by

$$\frac{\partial^2 U}{\partial \lambda^2} = -\frac{GM}{r} \sum_{n=o}^{N} \left(\frac{a}{r}\right)^n \sum_{m=o}^{n} m^2 \left[C_{nm}\cos m\lambda + S_{nm}\sin m\lambda\right] P_{nm}(\sin\psi),$$

$$\frac{\partial^2 U}{\partial \lambda \partial \psi} = \frac{GM}{r} \sum_{n=o}^{N} \left(\frac{a}{r}\right)^n \sum_{m=o}^{n} m \left[-C_{nm}\sin m\lambda + S_{nm}\sin m\lambda\right] \frac{dP_{nm}(\sin\psi)}{d\psi},$$

$$\frac{\partial^2 U}{\partial \lambda \partial r} = -\frac{GM}{r^2} \sum_{n=o}^{N} (n+1) \left(\frac{a}{r}\right)^n \sum_{m=o}^{n} \left[-C_{nm}\sin m\lambda + S_{nm}\cos m\lambda\right] P_{nm}(\sin\psi),$$

$$\frac{\partial^2 U}{\partial \psi^2} = \frac{GM}{r} \sum_{n=o}^{N} \left(\frac{a}{r}\right)^n \sum_{m=o}^{n} \left[C_{nm}\cos m\lambda + S_{nm}\sin m\lambda\right] \frac{d^2 P_{nm}(\sin\psi)}{d\psi^2},$$

$$\frac{\partial^2 U}{\partial \psi \partial r} = -\frac{GM}{r^2} \sum_{n=o}^{N} (n+1) \left(\frac{a}{r}\right)^n \sum_{m=o}^{n} \left[C_{nm}\cos m\lambda + S_{nm}\sin m\lambda\right] \frac{dP_{nm}(\sin\psi)}{d\psi},$$

$$\frac{\partial^2 U}{\partial r^2} = \frac{GM}{r^3} \sum_{n=o}^{N} (n+1)(n+2) \left(\frac{a}{r}\right)^n \sum_{m=o}^{n} \left[C_{nm}\cos m\lambda + S_{nm}\sin m\lambda\right] P_{nm}(\sin\psi).$$

The Legendre functions are evaluated recursively. Initial values are given by

$$P_{oo}(\sin\psi) = 1,$$

$$P_{1o}(\sin\psi) = \sin\psi,$$

$$P_{11}(\sin\psi) = \cos\psi.$$

For the Legendre polynomials, the recursion relation is

$$P_n(\sin\psi) = [(2n-1)\sin\psi\, P_{n-1}(\sin\psi) - (n-1)P_{n-2}(\sin\psi)]/n \qquad n\geq 2.$$

For the associated functions,

$$P_{nm}(\sin\psi) = P_{n-2,m}(\sin\psi) + (2n-1)\cos\psi\, P_{n-1,m-1}(\sin\psi) \qquad \begin{matrix} n\geq 2, \\ m\geq 1. \end{matrix}$$

where the first term on the right side is defined to be zero for $m>n-2$.

For the derivatives,

$$\frac{dP_{nm}(\sin\psi)}{d\psi} = P_{n,m+1}(\sin\psi) - m\tan\psi\, P_{nm}(\sin\psi)$$

where the first term on the right side is defined to be zero for $m>n-1$.

The second derivative $\frac{d^2P_{nm}}{d\psi^2}$ may be evaluated in terms of P_{nm} and $\frac{dP_{nm}}{d\psi}$ by using the differential equation satisfied by the Legendre associated functions,

$$\frac{d^2P_{nm}}{d\psi^2} = \tan\psi\,\frac{dP_{nm}}{d\psi} - \left[n(n+1) - \frac{m^2}{\cos^2\psi}\right] P_{nm}$$

The computational form of F is given by (4.30), and the computational form of G by (4.33). The full expression for G also contains terms of the form $\frac{\partial^2 T}{\partial x_i\,\partial \bar{x}_j}$. However, since the disturbing potential T is linear in the φ_K, these terms are also linear in the φ_K. And since the φ_K are all zero in the gravity field that generates the nominal orbit, all of these terms vanish.

In the generation of the "true" orbit used to simulate actual data, a term $\nabla_i T = \frac{\partial T}{\partial x_i}$ is added to the components of force obtained by (4.30). This is given in terms of earth fixed rectangular coordinates by (3.19). The computational form in inertial coordinates is then

$$\nabla T = \frac{\partial T}{\partial X} = -R_3(-\theta)\sum_k \varphi_k \Delta S_k \frac{1}{4}\sum_{i=1}^{4}\frac{1}{\ell_{ik}^3}(R_3(\theta)X - X_{ik}). \qquad (4.34)$$

The matrix $R_3(\theta)$ is the orthogonal rotation matrix which effects a rotation of the coordinate system around its third axis through an angle θ. The coordinates X are those of the satellite in the inertial coordinate system,

$R_3(\theta)X$ are the earth fixed coordinates of the satellite, and X_{ik} are the earth fixed coordinates of the center of the i th subblock of the k th block. The distance ℓ is computed from the difference in the earth fixed coordinates of the satellite and the subblock.

Since the normal potential U does not contain the density parameters φ_k, the matrix H may be obtained immediately by differentiating equation (4.34),

$$H = \frac{\partial f(X)}{\partial \varphi_k} = \frac{\partial^2 T}{\partial X \partial \varphi_k} = -R_3(-\theta)\,\Delta S_k \frac{1}{4} \sum_{i=1}^{4} \frac{1}{\ell_{ik}^3} (R_3(\theta)X - X_{ik}). \qquad (4.35)$$

4.24 Forming the Observation Equation. The total system to be integrated simultaneously by the predictor-corrector process consists of 84 differential equations: six elements of the orbit and 36 elements of the transition matrix for each of the two satellites. The initial condition for each of these 84 quantities is set up, the Ψ matrix for each satellite is initialized to zero, and the numerical integration is begun with a step size equal to that used in generating the simulated data. The functions to be integrated are given by equation (4.4) for the trajectory and (4.14) for the transition matrix for each satellite. After each step of the integration the H and Ψ' matrices are computed and the Ψ matrix is updated to the current epoch, using equation (4.26) if the total number of steps is odd and (4.28) if this number is even.

The file of simulated observations is then examined for a simulated observation at the current epoch. Since the simulated observations are formed at intervals that are multiples of the initial step size, the epoch of an observation will always be reached after an integral number of steps during the integration of the nominal orbits and transition matrices. When real data is processed, this will not generally happen, and an algorithm designed to process real data will need to invoke some kind of interpolation procedure in order to obtain values of Y_N, Φ, and Ψ for each orbit corresponding to the time of the observation.

If no simulated observation is found, the integration proceeds on to the

next integration step. If an observation is found, then the W matrix for each satellite is evaluated by equation (4.24) in the form $W = \cdot \Phi \Psi$. For a range rate observation, a computed observable $\dot{\rho}_N$ is computed by evaluating (4.5) with the elements of the nominal orbits $Y_N(t)$. This is subtracted from the "observed" $\dot{\rho}$ to form the constant term in the observation equations. The partial derivatives of the range rate with respect to the elements of each orbit are computed by (4.9). The coefficients of the observation equations are formed by combining these with the elements of the parameter sensitivity matrix in equation (4.8) and with the elements of the transition matrix in (4.7). The coefficients are arranged in a row matrix so that the observation equation can be written

$$A\,\delta\varphi_k + B_1\,\delta Y_{O1} + B_2\,\delta Y_{O2} = \dot{\rho} - \dot{\rho}_N \qquad (4.36)$$

where

$$A = \left[\frac{\partial \dot{\rho}}{\partial \varphi_k}\right],$$

$$B_1 = \left[\frac{\partial \dot{\rho}}{\partial Y_{O1}}\right] = \left[\frac{\partial \dot{\rho}}{\partial X_{O1}} \quad \frac{\partial \dot{\rho}}{\partial \dot{X}_{O1}}\right],$$

$$B_2 = \left[\frac{\partial \dot{\rho}}{\partial Y_{O2}}\right] = \left[\frac{\partial \dot{\rho}}{\partial X_{O2}} \quad \frac{\partial \dot{\rho}}{\partial \dot{X}_{O2}}\right].$$

If an observation of the position of one of the satellites is found at the current epoch, then three observation equations are formed, one for each coordinate. Assuming that the first satellite is observed, the three observation equations are

$$W_{11}\,\delta\varphi_k + \Phi_{11}\,\delta Y_{O1} = X_1 - X_{N1} \qquad (4.37)$$

The matrices W_{11} and Φ_{11} are obtained from the first three rows of the W and Φ matrices for the first satellite. The coefficient of δY_{O2} is zero. If the second satellite is observed, all quantities refer to the second satellite and the coefficient of δY_{O1} is zero.

All observation equations of the form (4.36) or (4.37) are recorded for later adjustment in a simulated solution.

4.3 The Energy Integral.

A useful check on the numerical integration of the orbits is provided by the energy integral. The classical form of the energy integral for time independent potentials is

$$\tfrac{1}{2}v^2 - V = \text{constant}$$

where v denotes the velocity and V the potential. However, this equation can be used only if there are no longitude dependent terms in the geopotential, since longitude dependent terms cause the potential to vary with time; i.e., although the potential is a function only of position in earth fixed coordinates, the earth rotates with time so that in inertial coordinates the potential is a function of both position and time. In this case, the classical integral is not valid, and the appropriate tool is the Jacobi integral [Hotine and Morrison, 1969].

The Jacobi integral is simply derived as follows. Again, let V denote the gravitational potential and let W denote the potential of gravity, obtained from the gravitational potential by the addition of the potential of centrifugal force,

$$W = V + \tfrac{1}{2}\omega^2 (x^2 + y^2)$$

where ω is the angular velocity of the earth. If X' denotes the time derivative of the position vector with respect to a basis that is rotating with the earth, and $\dot{X}$ denotes the time derivative of the same vector with respect to a non-rotating basis that is instantaneously coincident with the rotating basis, then

$$\dot{X} = X' + \Omega \times X$$

where Ω is the angular velocity vector of the earth. In either celestial or earth fixed coordinates

$$\Omega = \begin{pmatrix} 0 \\ 0 \\ \omega \end{pmatrix}$$

If Ω is considered to be constant (which is very nearly true), the relation between the accelerations is

$$\ddot{X} = X'' + 2\Omega \times X' + \Omega \times (\Omega \times X).$$

By Newton's third law, the acceleration in inertial coordinates $\ddot{X}$ is equal to the force per unit mass, which is the gradient of the gravitational potential V, so that

$$X'' = \nabla V - 2\Omega \times X' - \Omega \times (\Omega \times X).$$

The second term on the right is the Coriolis force and the third term is the centrifugal force. The usual form of the equations of motion in earth fixed Cartesian coordinates is obtained by moving the Coriolis term to the left hand side of the equation and resolving this vector equation into its components in earth fixed coordinates, taking account of the definition of Ω:

$$x'' - 2\omega y' = \frac{\partial V}{\partial x} + \omega^2 x,$$

$$y'' + 2\omega x' = \frac{\partial V}{\partial y} + \omega^2 y,$$

$$z'' = \frac{\partial V}{\partial z}.$$

This set of equations may also be written in terms of the potential of gravity W as

$$X'' + 2\Omega \times X' = \nabla W.$$

The Jacobi integral is obtained by taking the inner product of this equation with X',

$$X'' \cdot X' = \nabla W \cdot \dot{X}'.$$

Since W is a function only of position in earth fixed coordinates, $\nabla W \cdot X'$ is the total time derivative of W. Thus

$$X'' \cdot X' = \tfrac{1}{2}\frac{d}{dt}(X' \cdot X') = \nabla W \cdot X' = \frac{dW}{dt}$$

Integrating,

$$\tfrac{1}{2}X' \cdot X' = W + C.$$

Substituting $X' = \dot{X} - \Omega \times X$, this becomes

$$\tfrac{1}{2}\dot{X} \cdot \dot{X} - (\Omega \times X) \cdot \dot{X} + \tfrac{1}{2}(\Omega \times X)^2 = W + C.$$

Writing this in terms of the inertial position and velocity components yields

$$\tfrac{1}{2}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - \omega(x\dot{y} - y\dot{x}) + \tfrac{1}{2}\omega^2 (x^2 + y^2) = W + C,$$

or

$$\tfrac{1}{2}v^2 - V - \omega(x\dot{y} - y\dot{x}) = \text{constant} \qquad (4.38)$$

which is the Jacobi integral.

The quantity (4.38) is computed at each step of the integration of the simulated "true" orbit and the nominal orbit for each satellite. Although this quantity is different for different orbits, its constancy for any one orbit serves as a useful check on the accuracy of the numerical integration process.

4.4 Solution of the Observation Equations.

The observation equations are collected together and solved in a conventional least squares adjustment. Since the simulated observations do not contain any observational errors, it should be possible to obtain a solution that satisfies all the equations, at least to the extent that the linearized form of the observation equations is valid. Since the parameters used to generate the simulated observations are known, the differences between the recovered parameters and these "true" parameters provide a measure of the numerical error of the algorithm.

All of observation equations arising from a single pass of the two satellites may be collected together and written as a single matrix equation

$$\overline{A}_j \, \delta\varphi_K + \overline{B}_{1j} \, \delta Y_{1j} + \overline{B}_{2j} \, \delta Y_{2j} = L_j \quad . \tag{4.39}$$

where the subscript j is attached to denote the jth pair of orbits. The gravity field unknowns are common to all orbits, but a new set of orbit unknowns is introduced for each pass of each satellite. Weights are introduced for each observation equation by inverting the assumed variance for the corresponding observation. These are arranged in a diagonal weight matrix P_j for each pair of orbits. The normal equations are written in partitioned form as

$$\begin{pmatrix} \dot{N} & \overline{N}_1 & \overline{N}_2 & \cdots & \overline{N}_n \\ \overline{N}_1^T & \ddot{N}_1 & 0 & \cdots & 0 \\ \overline{N}_n^T & 0 & \ddot{N}_2 & \cdots & 0 \\ \vdots & \vdots & & & \\ \overline{N}_n^T & 0 & \cdots & \cdots & \ddot{N}_n \end{pmatrix} \begin{pmatrix} \delta\varphi_K \\ \delta Z_1 \\ \cdot \\ \cdot \\ \delta Z_n \end{pmatrix} = \begin{pmatrix} \dot{K} \\ \ddot{K}_1 \\ \cdot \\ \cdot \\ \ddot{K}_n \end{pmatrix} \tag{4.40}$$

where

$$\dot{N} = \sum_j \overline{A}_j^T \, P_j \, \overline{A}_j \, ,$$

$$\overline{N}_j = \overline{A}_j^T \, P_j \, \overline{B}_j$$

$$\ddot{N}_j = \bar{B}_j^T P_j \bar{B}_j ,$$

$$\dot{K} = \sum_j \bar{A}_j^T P_j L_j ,$$

$$\ddot{K}_j = \bar{B}_j^T P_j L_j ,$$

and

$$\bar{B}_j = (\bar{B}_{1j}\ \bar{B}_{2j}) , \qquad \delta Z_j = \begin{pmatrix} \delta Y_{1j} \\ \delta Y_{2j} \end{pmatrix} .$$

Since the orbit unknowns are not of any particular interest, they may be eliminated from the system of equations. The resulting reduced normal equations are written

$$\tilde{N}\, \delta\varphi_K = \tilde{K} \tag{4.41}$$

where

$$\tilde{N} = \dot{N} - \sum_j \bar{N}_j\, \ddot{N}_j^{-1}\, \bar{N}_j^T$$

$$\tilde{K} = \dot{K} - \sum_j \bar{N}_j\, \ddot{N}_j^{-1} \ddot{K}_j$$

This equation is solved for the gravity field parameters $\delta\varphi_K$. The inverse of the coefficient matrix $\tilde{N}$ is the covariance matrix of these parameters, and it furnishes the uncertainties of the parameters and the correlations between them. From these statistics, from the conditioning of the normal equation coefficient matrix, and from the numerical error of the solution, it is possible to make a qualitative judgment as to whether observations of the type used can successfully resolve the components of the gravity field characterized by the φ_K.

5. THE SIMULATED DATA AND SOLUTIONS

The algorithm described in the previous chapter was coded for machine computation in FORTRAN IV language. Several series of simulated solutions were performed on the IBM 360/75 computer of The Ohio State University. All computations were performed in double precision arithmetic.

The normal gravity field used for all computations was the 1969 Smithsonian Standard Earth [Gaposchkin and Lambeck, 1969] field truncated at degree and order (12, 12). This field represents most of the information that may be obtained from ground to satellite tracking data alone, without the addition of surface gravimetry data [Gaposchkin and Lambeck, 1969; Gaposchkin, 1970]. The reference ellipsoid and other parameters associated with this gravity field were also used in all appropriate computations. A contour map of the geoid was produced using the full list of harmonic coefficients, and this map served to define the geoid height, when needed.

The fictitious surface layer was spread on the surface defined by $U = W_o$, which closely approximates the geoid. The topographic heights of land areas were not considered, since they were not pertinent to the purpose of the study. Block sizes of $5^\circ \times 5^\circ$, $2^\circ \times 2^\circ$, and $1^\circ \times 1^\circ$ were used to describe the density of the surface layer. None of the solutions considered a global surface layer spread over the entire earth. The largest area considered covered roughly the area of the contiguous United States and the Caribbean, described by 92 $5^\circ \times 5^\circ$ blocks. Smaller areas were used when solutions for smaller block sizes were considered, so that no solution involved more than 100 parameters describing the gravity field. The density of the surface layer outside the area being considered was assumed to be zero. No allowance was made for these neglected areas in simulated solutions, and the possibility that the effect of such areas might bias the solution was not considered.

A series of passes over the area being considered was generated for each of the assumed gravity fields. Satellite altitudes of 700 km, 300 km, 200 km, and 100 km were used, the choice depending on the block size used to describe the gravity field. In each case, satellite to satellite range rate and satellite position observations were simulated, observation equations were formed, and a solution for the unknown parameters describing the gravity field was performed.

Initial step sizes of 32 and 16 seconds were used for the integration of the orbits and the partial derivatives, depending on the density of observations desired. The tolerance and the weights used to control the local truncation error during the integration process were selected to provide an estimated global error of less than 1 meter in position and 1×10^{-4} meter per second in velocity. This corresponds to a numerical accuracy of about eight significant decimal digits in all quantities. Examination of the constancy of the energy integral discussed in Section 4.3 indicated that accuracy in this many significant digits was maintained. For passes at an altitude of 700 km, a step size of 32 seconds was sufficient to maintain this accuracy. For passes at lower altitudes, the integration module selected a step size of 16 seconds, regardless of the initial step size.

The weights assigned to the different kinds of observation equations were varied, although the weight of the satellite to satellite range rate observation was most often formed from an observational standard deviation of 0.05 mm/sec. The predicted range of accuracies for this type of observation is 0.03 - 0.05 mm/sec [Kaula, 1969], so that the figure used is slightly on the conservative side.

Several tests were performed to assure that the linearized observation equations were valid for the range of values of orbit unknowns under consideration. Since the "true" set of values of the unknowns was known, these values could be substituted into the linearized observation equations. If these values satisfied the linearized observation equation, at least within

the uncertainty assigned to that type of observation, then the equation was judged to be satisfactorily linear for that set of values. The tests indicated that the linearized observation equations were valid for mean densities of the surface layer of 10 to 15 mgal, variations in initial epoch position elements of several hundred meters, and variations in initial velocity elements of several tenths of a meter per second. This indicated the validity of integrating the transition matrix with the same step size used to integrate the orbit, as well as the validity of the algorithm used to generate the parameter sensitivity matrix. It also indicated that, with the expected values of the unknowns, a solution can be reached in a single iteration.

The algorithm used to solve the reduced normal equations was the LU algorithm [Conte, 1965, p. 178], modified specifically for symmetric matrices [Uotila, 1967, p. 27]. Computer storage space was reserved only for the upper triangular part of the matrix of coefficients. With this storage restriction, symmetry must be maintained at each stage of the reduction, so that pivoting can take place only on the main diagonal. I.e., whenever two rows are interchanged, the corresponding columns must also be interchanged to maintain symmetry. It was noticed that if the matrix is positive definite, then the ratio of the maximum to the minimum pivot element provides a good indicator of its condition. This number can be shown to be directly related to the M-number of Turing [Faddeev and Faddeeva, 1963, p. 125]. This condition number was used in conjunction with the covariance matrix of recovered parameters to judge whether a given set of observations was capable of resolving the mean densities of the surface layer in a given set of blocks.

The potential represented by the surface layer can also be represented by spherical harmonic coefficients by equation (3.8) in the form

$$\begin{Bmatrix} \Delta C_{nm} \\ \Delta S_{nm} \end{Bmatrix} = \frac{2 - \delta_{0m}}{GM} \frac{(n-m)!}{(n+m)!} \sum_k \varphi_k \sum_{i=1}^{4} \left(\frac{\rho_{ik}}{a} \right)^n \begin{Bmatrix} \cos m\lambda_{ik} \\ \sin m\lambda_{ik} \end{Bmatrix} \int P_{nm}(\sin \psi_{ik}) \frac{\Delta S_k}{4}$$

where (ρ_{ik} , ψ_{ik} , λ_{ik}) are the spherical coordinates of the ith sub-block of the kth block. Were the values of the density computed in blocks covering the whole earth, we would expect the ΔC_{nm} , ΔS_{nm} to all be zero for (n, m) less than (12, 12), since the low order components of the gravity field are contained in the modified normal potential and not in the modified disturbing potential. However, this is not true when only a portion of the earth's surface is considered. Nevertheless, values of ΔC_{nm} and ΔS_{nm} may be computed from assumed values of the φ_k , and these equations may then be used to impose linear constraints on the φ_k during the simulated solution. This procedure may be interpreted as expressing the condition that the values of the potential coefficients below (12, 12) are not to be changed. This condition will insure that the surface layer within and outside of the area under consideration will not give rise to any new terms in the disturbing potential of degree 12 or less.

Since the potential coefficients of degree and order (12, 12) and lower were assumed to be perfectly known, it would have been possible to impose $13^2 = 169$ linear constraints on the values of the mean densities. However, the largest solution involved the unknown mean density in only 100 blocks. Since the imposition of the larger set of constraints would have completely overridden the observational data, and would have masked the ability of the observations to separate the effects of the densities of neighboring blocks, these constraints were not exercised.

5.1 Preparation of Assumed Gravity Fields.

5.11 Method of Preparation. Although any assumed gravity field would serve equally well for a simulation study, it was decided to use the gravity field as it is actually known from terrestrial gravimetry, since this would best represent the magnitude of the irregularities we hoped to detect, and would also demonstrate the conversions involved in transforming between different representations of the gravity field. The terrestrial field is usually represented by mean free air gravity anomalies referred to the

International Gravity Formula, while the representation desired was the mean densities of blocks of a surface layer; therefore, a transformation from gravity anomalies to surface layer densities was necessary.

The anomalies are assumed to be referred to an ellipsoid of flattening 1/297 inherent in the International Gravity Formula. Since the 1969 Smithsonian Standard Earth is used to represent the normal field, it was first necessary to convert the gravity anomalies into this system. The parameters used in the 1969 Standard Earth are [Gaposchkin and Lambeck, 1970, p. 8, p. 49, p. 64]

$$\begin{aligned} GM &= 3.986013 \times 10^{14}\ m^3/sec^2, \\ a &= 6378155\ m, \\ 1/f &= 298.255. \end{aligned}$$

The first two parameters were used by Kozai in his solution for the zonal terms of the gravity field [Gaposchkin and Lambeck, 1970, p. 8], as well as in the solution for the tesseral terms. The value of GM was originally obtained by the Jet Propulsion Laboratory and was the value used to scale the earlier SAO C6 system [Lundquist and Veis, 1966]. Together with a value of the rate of rotation of the earth $\omega = 0.72921151467 \times 10^{-4}$ radian/sec (assumed known), the above values are sufficient to determine the other constants of the gravity field. It is stated [Gaposchkin and Lambeck, 1970, p. 49] that the value of $1/f = 298.255$ corresponds to Kozai's 1969 value of $J_2 = 1082.628 \times 10^{-6}$. However, this is not precisely true: using the other parameters the value $1/f = 298.255$ implies $J_2 = 1082.6392 \times 10^{-6}$, while $J_2 = 1082.628 \times 10^{-6}$ implies $1/f = 298.2565$. Thus, the ellipsoid $a = 6378155$ m, $1/f = 298.255$ is not precisely a mean earth ellipsoid. The value of $1/f = 298.255$ appears to have been obtained from Kozai's 1967 solution for zonal harmonics [Lundquist, 1967b].

The values of equatorial gravity and the potential on the geoid corresponding to the above parameters are

$$\gamma_e = 978.0291 \text{ gal},$$

$$W_o = 6263681.1 \text{ kgal meters}.$$

These values are well within the uncertainties of other recent determinations [Rapp, 1966], and thus may be assumed to define the actual earth. The equipotential ellipsoid a = 6378155 m, 1/f = 298.255 then has the same mass and rotational velocity as the actual earth, the same potential as the geoid, and is a volume ellipsoid [Mueller and Rockie, 1966]. The formula for normal gravity on this equipotential ellipsoid is

$$\gamma = 978.0291 \, (1 + .0053025 \sin^2\varphi - .00000585 \sin^2 2\varphi) \text{ gal}.$$

The value of equatorial gravity in the International Formula was obtained from an analysis of gravity values in the Potsdam system, which is known to differ significantly from the absolute system [Heiskanen and Moritz, 1967, p. 152]. Therefore, gravity measurements used to compute gravity anomalies with the International Formula should be in the Potsdam system; similarly, gravity measurements should be in the absolute system if the gravity formula above is used. The gravity anomaly in the International system is

$$\Delta g_I = g_o^P - \gamma_I$$

where g_o^P denotes measured gravity in the Potsdam system reduced to the geoid, and γ_I is normal gravity computed from the International Gravity Formula. In the SAO 1969 Standard Earth system, the gravity anomaly is

$$\Delta g_{SAO} = g_o^A - \gamma_{SAO} = \Delta g_I + (g_o^A - g_o^P) - (\gamma_{SAO} - \gamma_I)$$

where g_o^A is measured gravity reduced to the geoid in the absolute system, and γ_{SAO} is normal gravity in the SAO system computed by the gravity formula above. The term $(g_o^A - g_o^P)$ is the Potsdam correction, and the term $(\gamma_{SAO} - \gamma_I)$ is the difference of the two gravity formulae. Using a value of -13.7 mgal for the Potsdam correction yields

$$\Delta g_{SAO} = \Delta g_I + 6.2 - 13.7 \sin^2 \varphi \quad \text{(mgal)} \tag{5.1}$$

for the conversion of the gravity anomalies.

The mean density of a block of the surface layer may then be obtained from

$$\varphi = \frac{1}{2\pi} \left(\overline{\Delta g} + \frac{3\bar{\gamma}}{2R} \overline{N} \right) \tag{5.2}$$

where $\overline{\Delta g}$ is the mean gravity anomaly and $\overline{N}$ the mean geoid height in the block. A surface layer with this density generates the total disturbing potential in the conventional sense. However, the surface layer is intended to represent only that portion of the disturbing potential that corresponds to terms of degree higher than 12. Thus, it is necessary to remove that density distribution which gives rise to the difference between the potential represented by the spherical harmonic series of degree 12 and the potential of a level ellipsoid. The density distribution corresponding to the conventional disturbing potential is expressed in terms of spherical harmonic coefficients by equation (3.14)

$$\varphi = \sum_{n=o}^{\infty} \varphi_n = \frac{1}{4\pi R} \sum_{n=o}^{\infty} (2n+1)\, T_n \, .$$

Thus the modified density is

$$\varphi^* = \varphi - \frac{1}{4\pi R} \sum_{n=o}^{12} (2n+1)\, T_n \tag{5.3}$$

where the spherical harmonic coefficients of the 1969 Smithsonian Standard Earth are used to compute the T_n. The φ^* are the density parameters for which solutions were simulated. They are simply denoted by ω elsewhere.

5.12 The Gravity Fields. Four assumed gravity fields were prepared in the manner described in the previous section. Mean free air anomalies referred to the International Formula were converted to the SAO system by equation (5.1). The corresponding density was computed by (5.2) and the effect of the (12, 12) field was removed as in (5.3). The values of the geoid height used in (5.2) referred to the SAO ellipsoid a = 6378155.0 m, 1/f = 298.255. They were read from a contour map generated from the full set of SAO gravity coefficients. The four gravity fields are described below.

A. A set of 92 $5^\circ \times 5^\circ$ blocks in the vicinity of the United States and the Caribbean. This area was chosen because terrestrial gravity anomalies have been densely measured within it, and the actual roughness of the field is represented by their rapid variation. The mean free air anomalies for these blocks were taken from [Heiskanen and Moritz, 1967]. These are shown in Figure 5.1, and the geoid heights are shown in Figure 5.2. The corresponding values of the density parameters are shown in Figure 5.3.

Figure 5.4 shows the mean values of the density parameters after the (12, 12) portion of the gravity field was removed by equation (5.3). These are not a great deal smaller than the values in Figure 5.3, which somewhat contradicts the assumption that the dominant part of disturbing potential is given by terms of degree 12 and lower. On the other hand, this may also be interpreted as reflecting the roughness of the true gravity field, since the irregularities of $5^\circ \times 5^\circ$ blocks cannot be absorbed by harmonic analysis in which the shortest half-wave length used is 15°, corresponding to the 12th degree terms.

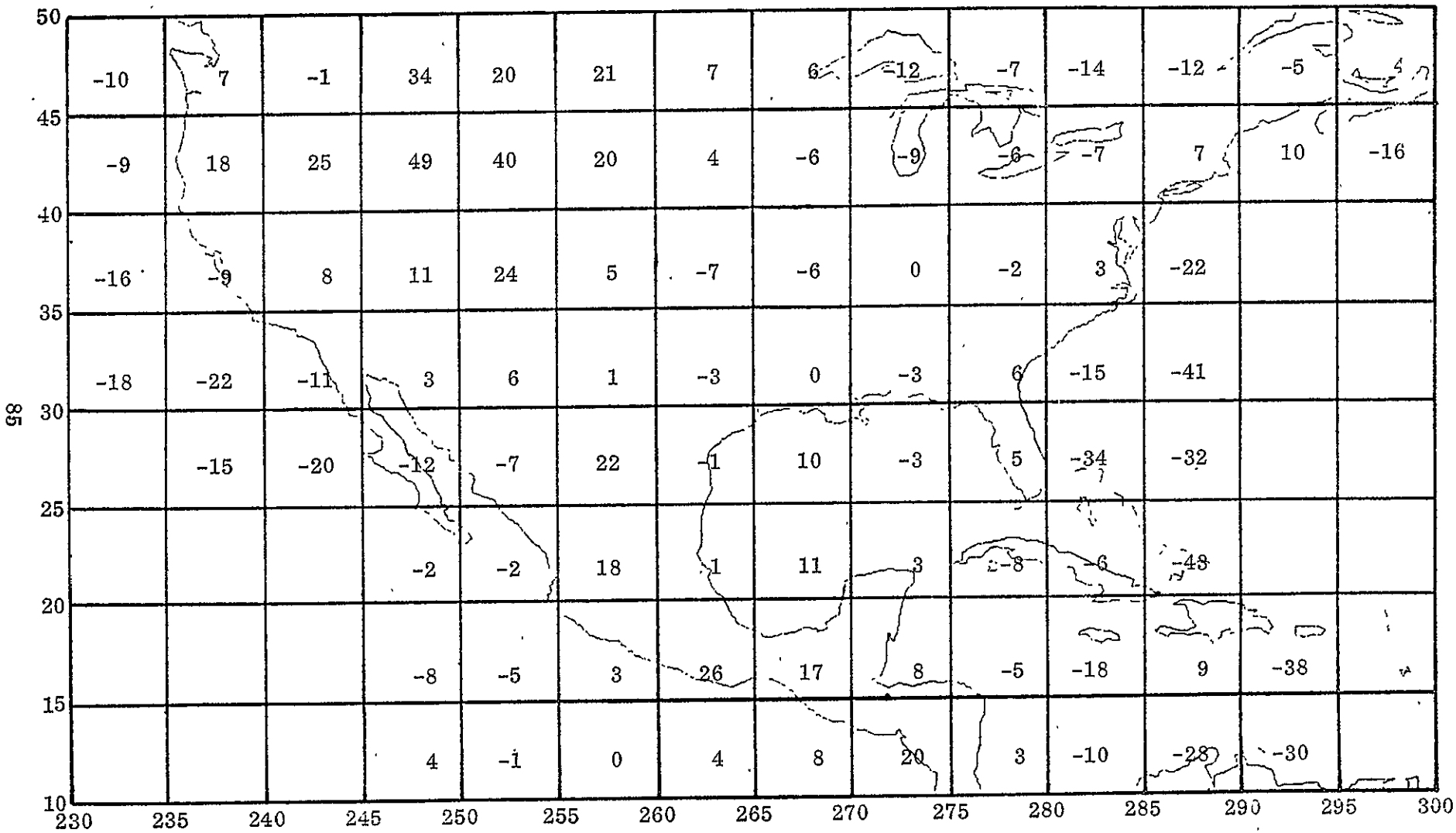

Fig. 5.1. Mean Free Air Anomalies in North America (mgal) (International Gravity Formula). [Heiskanen and Moritz, 1967, p. 153].

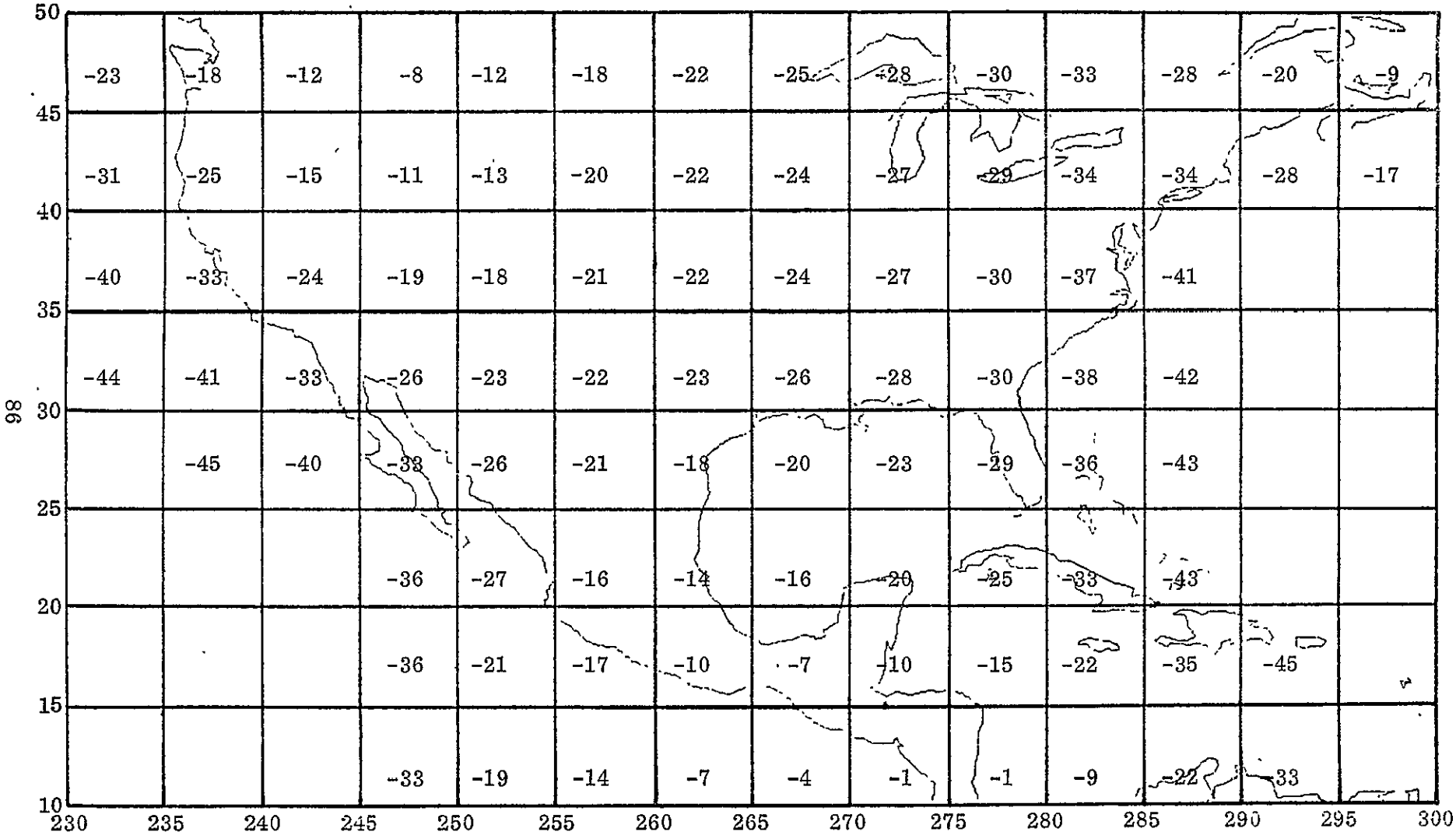

Fig. 5.2. Mean Geoid heights in Meters from the 1969 SAO Standard Earth Gravity Field [Gaposchkin and Lambeck, 1970]

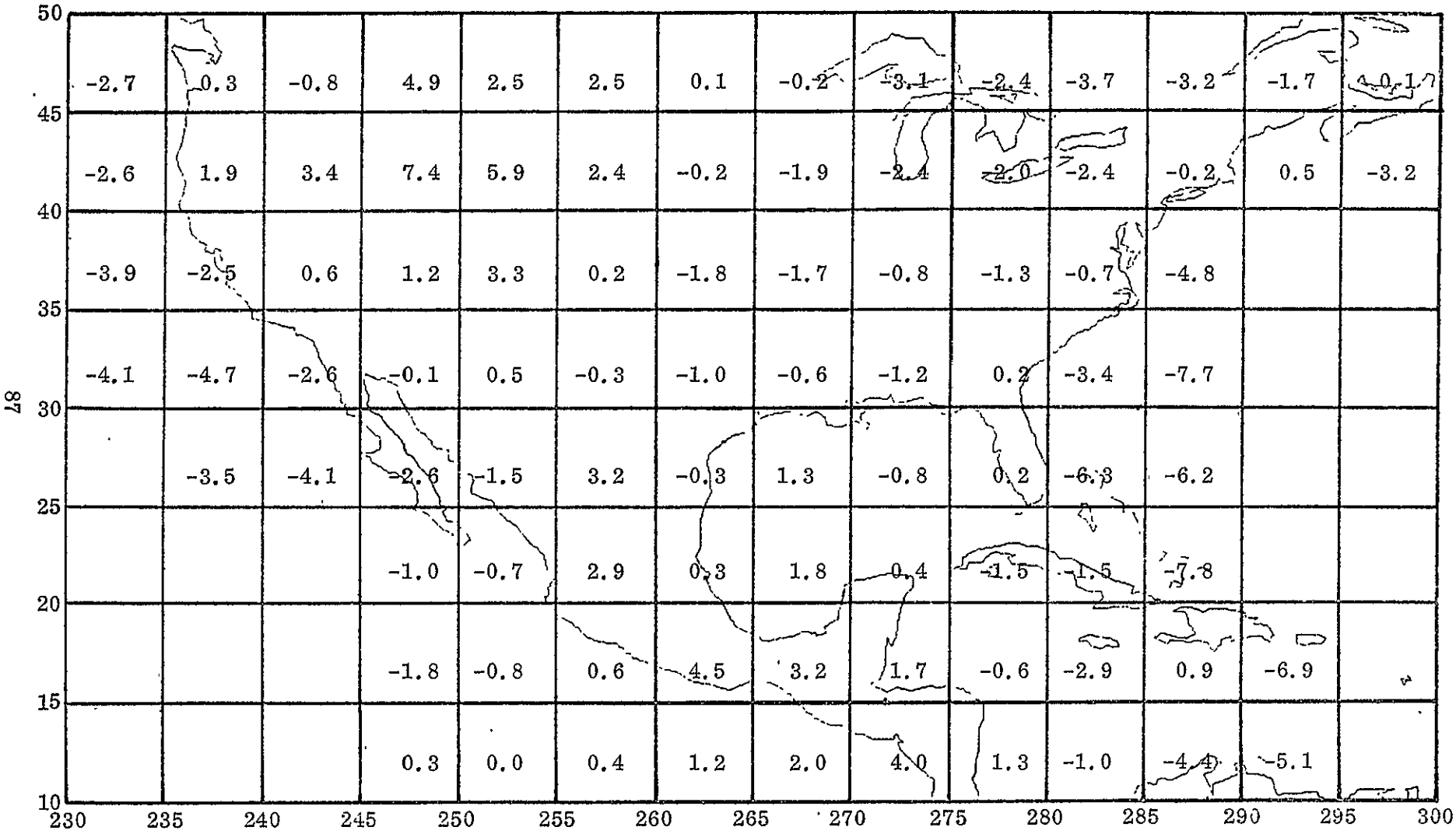

Fig. 5.3. $5^\circ \times 5^\circ$ Mean Values of the Density Parameter φ (mgal).

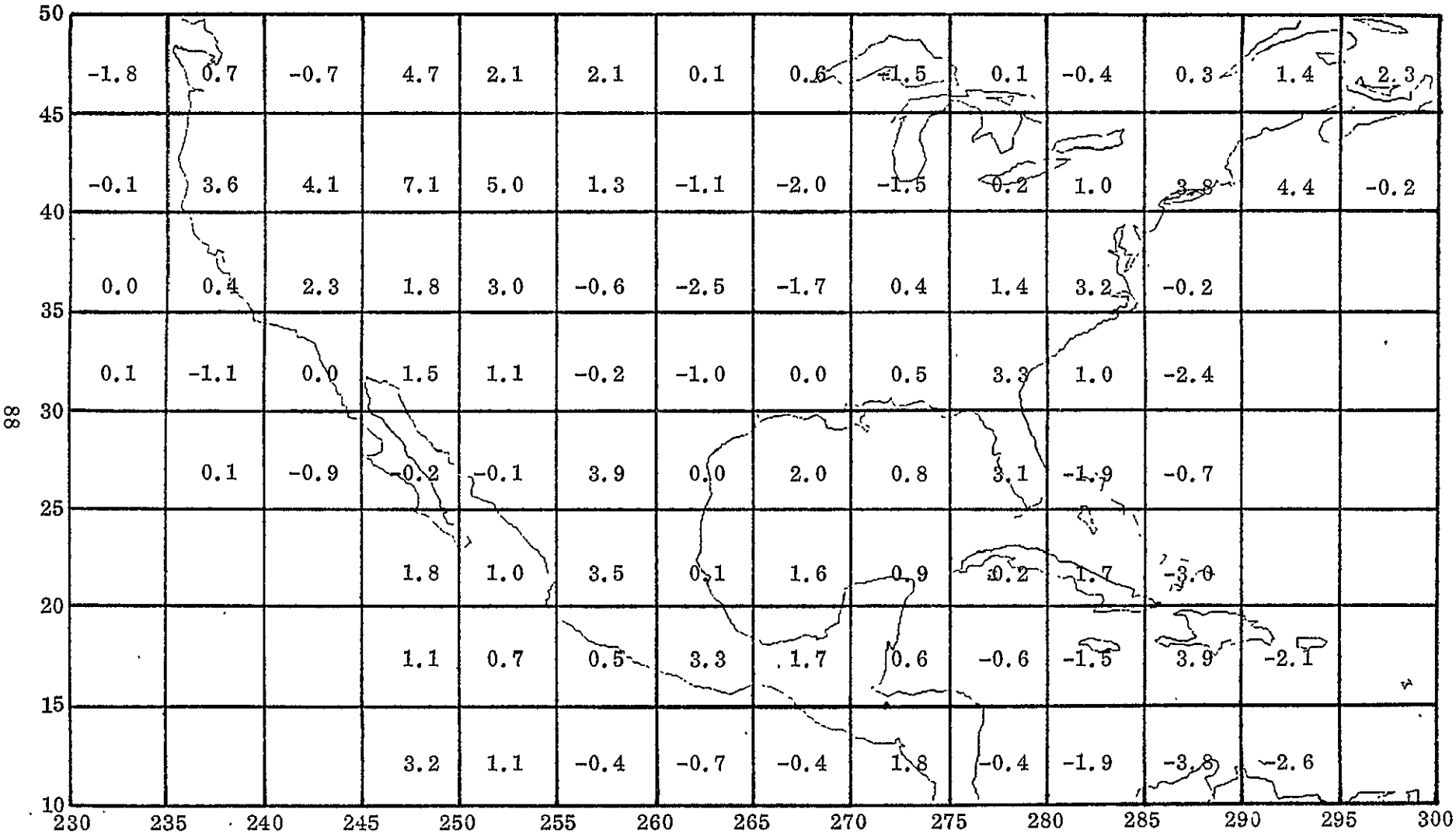

Fig. 5.4. $5^\circ \times 5^\circ$ Mean Values of the Density Parameters φ After Removal of the (12, 12) Portion of the Gravity Field (mgal).

B. A subset of set A, consisting of 24 $5^\circ \times 5^\circ$ blocks between 25° and 50° north latitude and 230° and 255° east longitude.

C. A set of 100 $1^\circ \times 1^\circ$ blocks between the limits $35^\circ - 45^\circ$ north latitude and $240^\circ - 250^\circ$ east longitude. Free air gravity anomalies for this area were taken from [Strange and Woollard, 1964] and are shown in Figure 5.5. The corresponding values of the density parameters, after removal of the (12, 12) portion of the gravity field, are shown in Figure 5.6. As expected, the gravity anomalies and density values show a much greater variation than in the case of the larger blocks.

D. The gravity anomalies and geoid heights used to generate set C were meaned together into 25 $2^\circ \times 2^\circ$ blocks. The free air gravity anomalies are shown in Figure 5.7; and the corresponding density parameters, after removal of the (12, 12) portion of the gravity field, are shown in Figure 5.8.

5.2 Sensitivity of the Range Rate to the Density of the Surface Layer.

The coefficients in the observation equations describe the sensitivity of the measured quantity to each of the unknown parameters. By plotting these coefficients on a map, one may identify the effect of each block on the range rate between the two satellites. This was done first for two satellites separated by 200 km in the same orbit at an altitude of 700 km. The partial derivatives with respect to the mean values of the density parameter in the 92 $5^\circ \times 5^\circ$ blocks (Set A) were computed for a point near the middle of the pass. These sensitivity coefficients are shown in Figure 5.9. They describe the effect on the range rate between the satellites of a block in which the density parameter is one mgal. The sensitivity is zero in the block beneath the two satellites; it reaches a positive maximum about 700 km in front of their position, and a negative extremum about 700 km behind their position. When these sensitivities are evaluated for several points along the same orbit, the most noticeable phenomenon is that the pattern shown in Figure 5.9 follows the

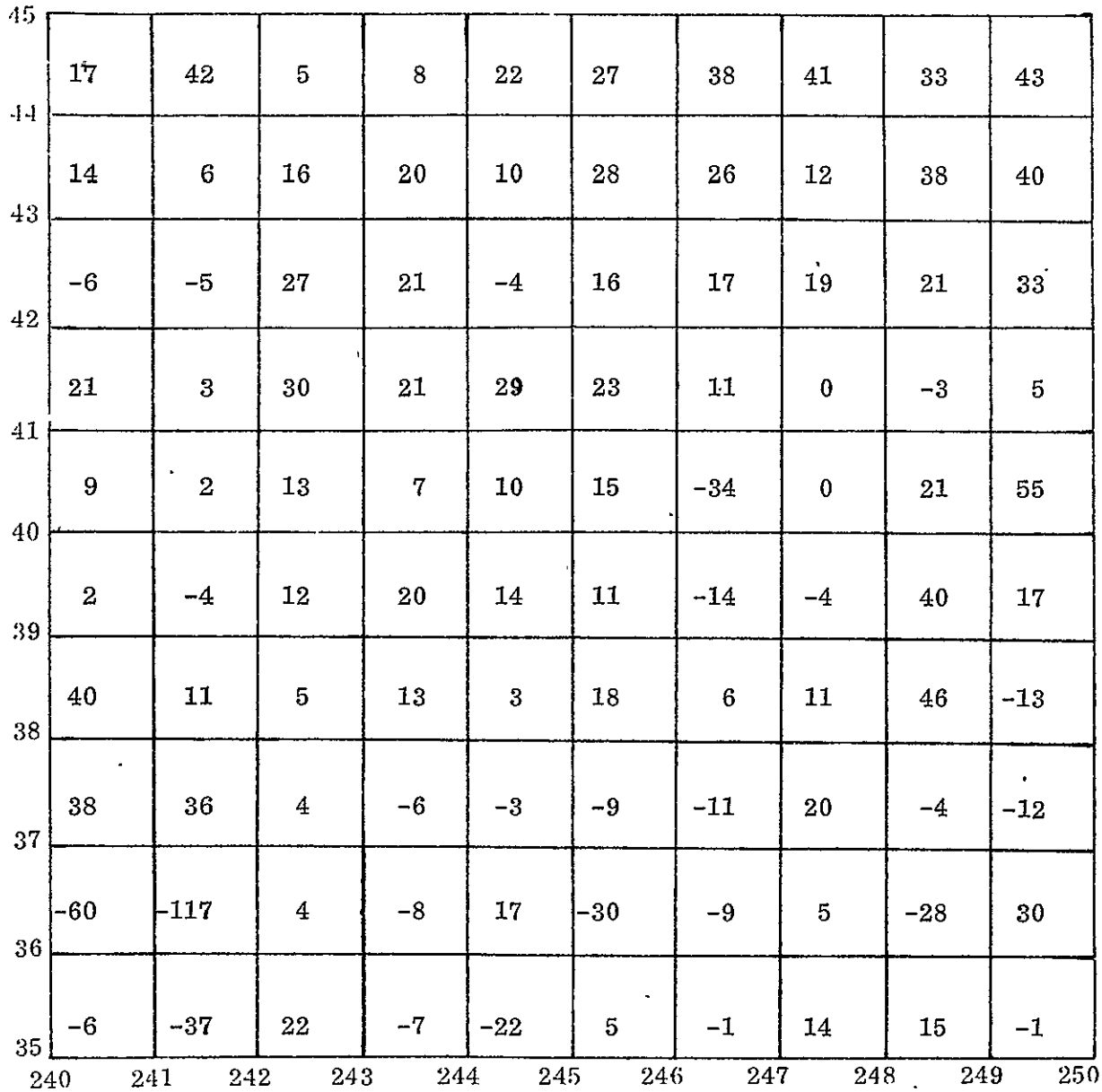

Fig. 5.5. Free Air Anomalies in $1^\circ \times 1^\circ$ Blocks (mgals), Referred to International Gravity Formula [Strange and Woollard, 1964].

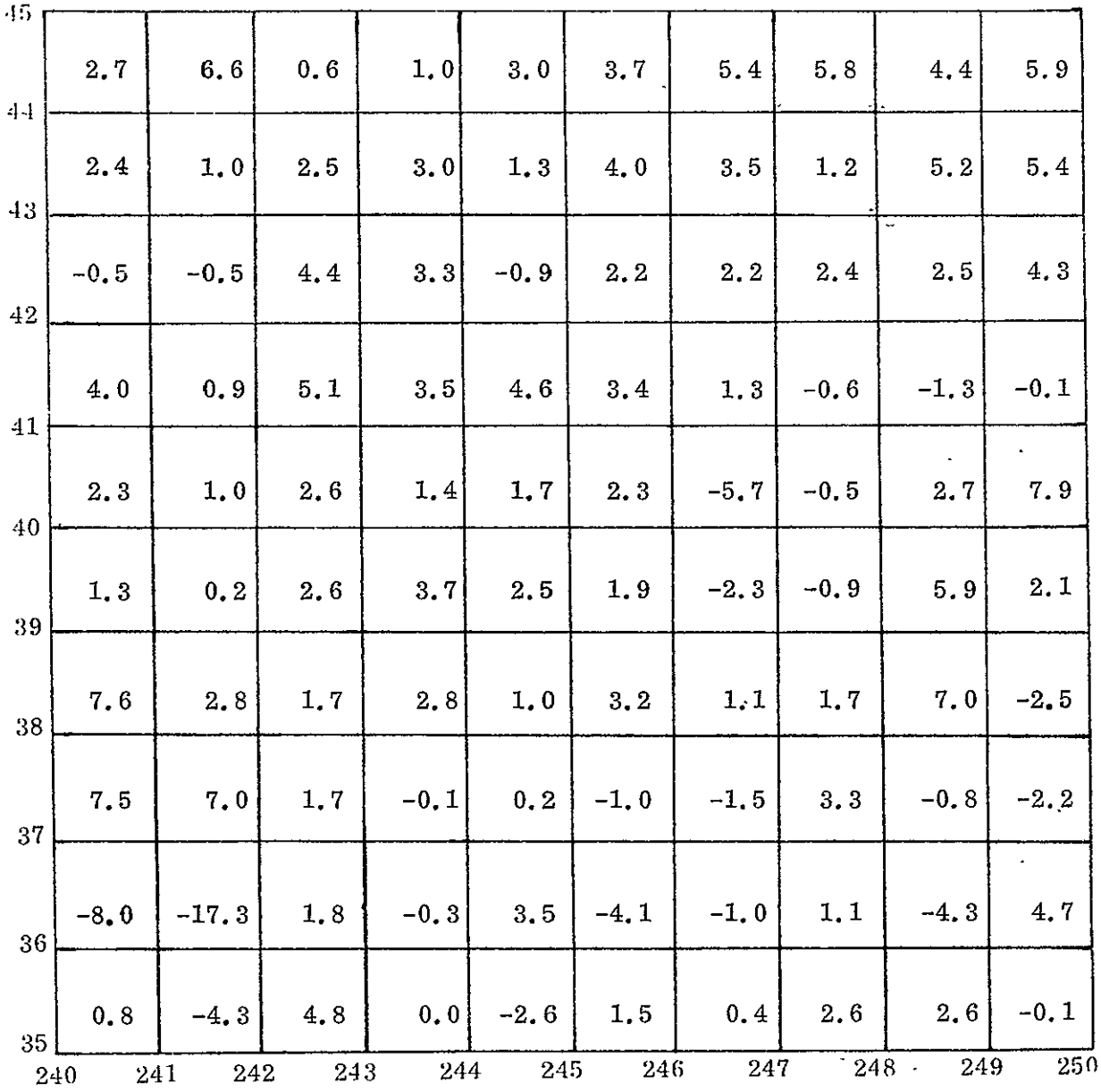

Fig. 5.6. Mean Values of the Density Parameter φ After Removal of the (12, 12) Portion of the Gravity Field (mgal).

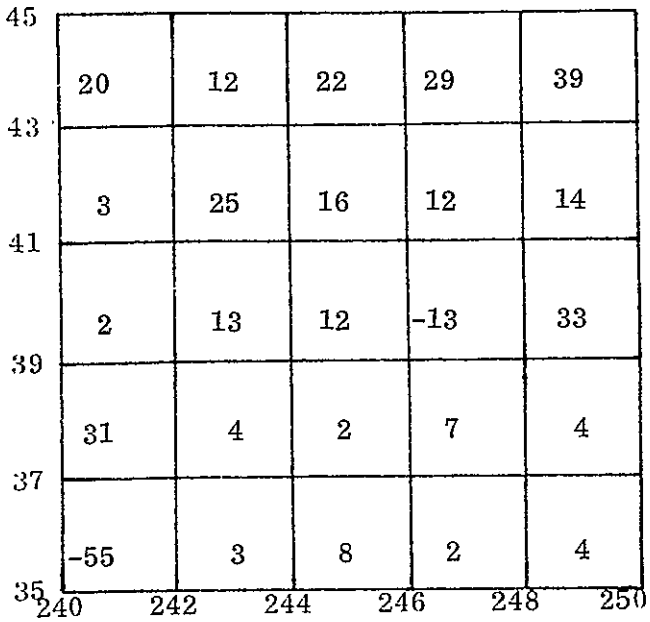

Fig. 5.7. Free Air Gravity Anomalies in $2^\circ \times 2^\circ$ Blocks Referred to the International Gravity Formula (mgals).

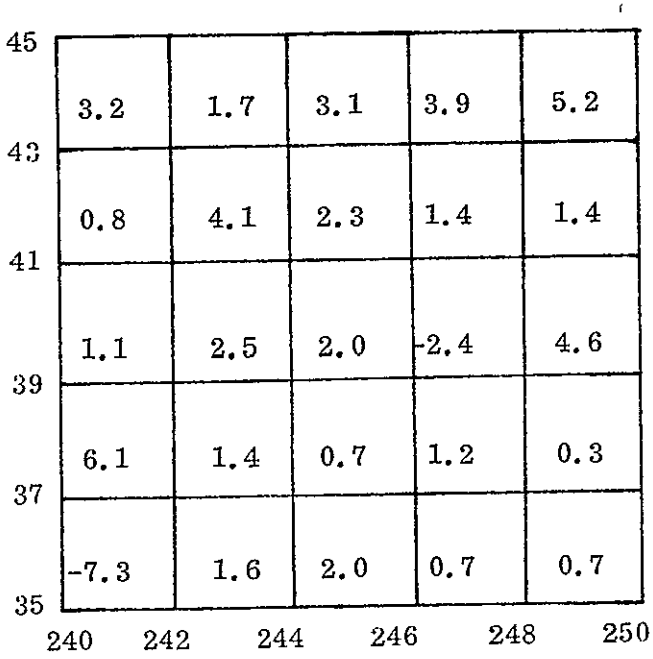

Fig. 5.8. Mean Values in $2^\circ \times 2^\circ$ Blocks of the Parameter φ After Removal of the (12, 12) Portion of the Gravity Field (mgals).

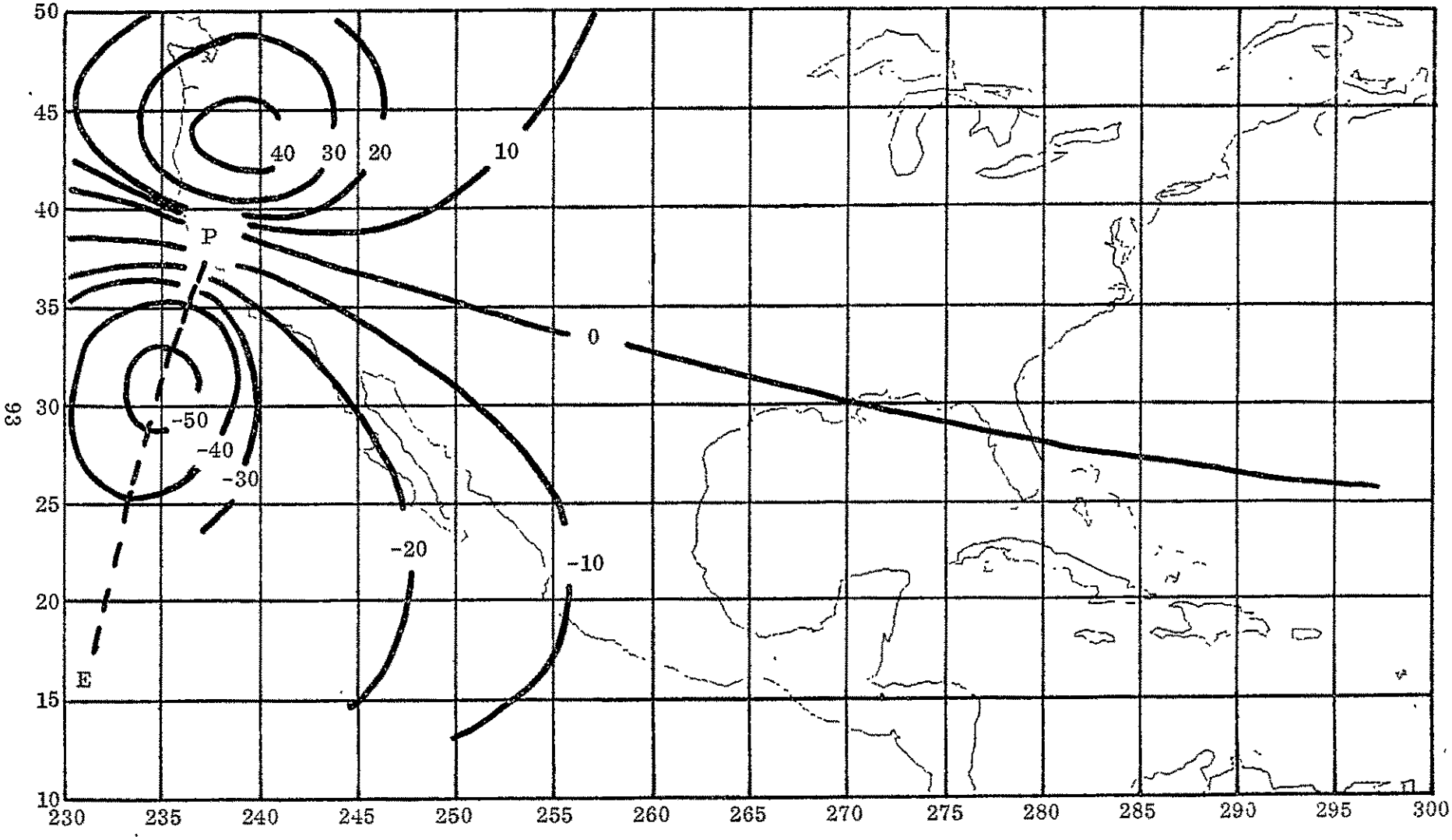

Fig. 5.9. Sensitivities of $\dot{\rho}$ in thousandths of a millimeter per second to φ in milligals for two satellites separated by 200 km in the same orbit 700 km high. E denotes the position at the initial epoch and P denotes the position for which the partial derivatives are evaluated.

satellites along the orbits. At whatever point the partial derivatives are evaluated, an area of maximum positive sensitivity is found a short distance in front of the position of the satellites, and an area of maximum negative sensitivity is found a short distance behind their position. Furthermore, the sensitivity is zero along a line drawn through the satellite's position and perpendicular to the ground path of the orbit. The sensitivities in back of this line are almost always negative and those in front of the line are positive. If the sensitivity coefficients are plotted as a function of time for a block lying on the ground path of the satellite, the typical sinusoidal signature discussed in Chapter 2 is obtained. The effect of the surface layer in blocks far from the ground path of the orbit remain small throughout the pass. Furthermore, the magnitude of the positive and negative maxima remain fairly constant at about 0.05 mm/sec. This means that at this altitude the effect on the range rate of a block on which the density parameter is one mgal and which lies on the ground path of the satellite is equal to the expected noise level in the measurement. Thus, the accuracy of the determination of the parameters from this altitude cannot be expected to be much better than one mgal, unless a great many observations are used. On the other hand, only a few blocks have significantly large sensitivities, and among these blocks sensitivities of the range rate to the values of the parameters in two neighboring blocks are significantly different. This means that this type of observation should be well able to separate the values of the density in neighboring blocks.

For purposes of comparison, partial derivatives were also evaluated for the configuration of one low satellite tracked by a high geostationary satellite. The low satellite was in the same 700 km high orbit used for both satellites in the previous case, and the $5^\circ \times 5^\circ$ blocks were again used. The sensitivity coefficients for this case are much larger, as shown in Figure 5.10. In the case of two satellites in the same low orbit, the far away blocks affect both satellites in approximately the same way, and

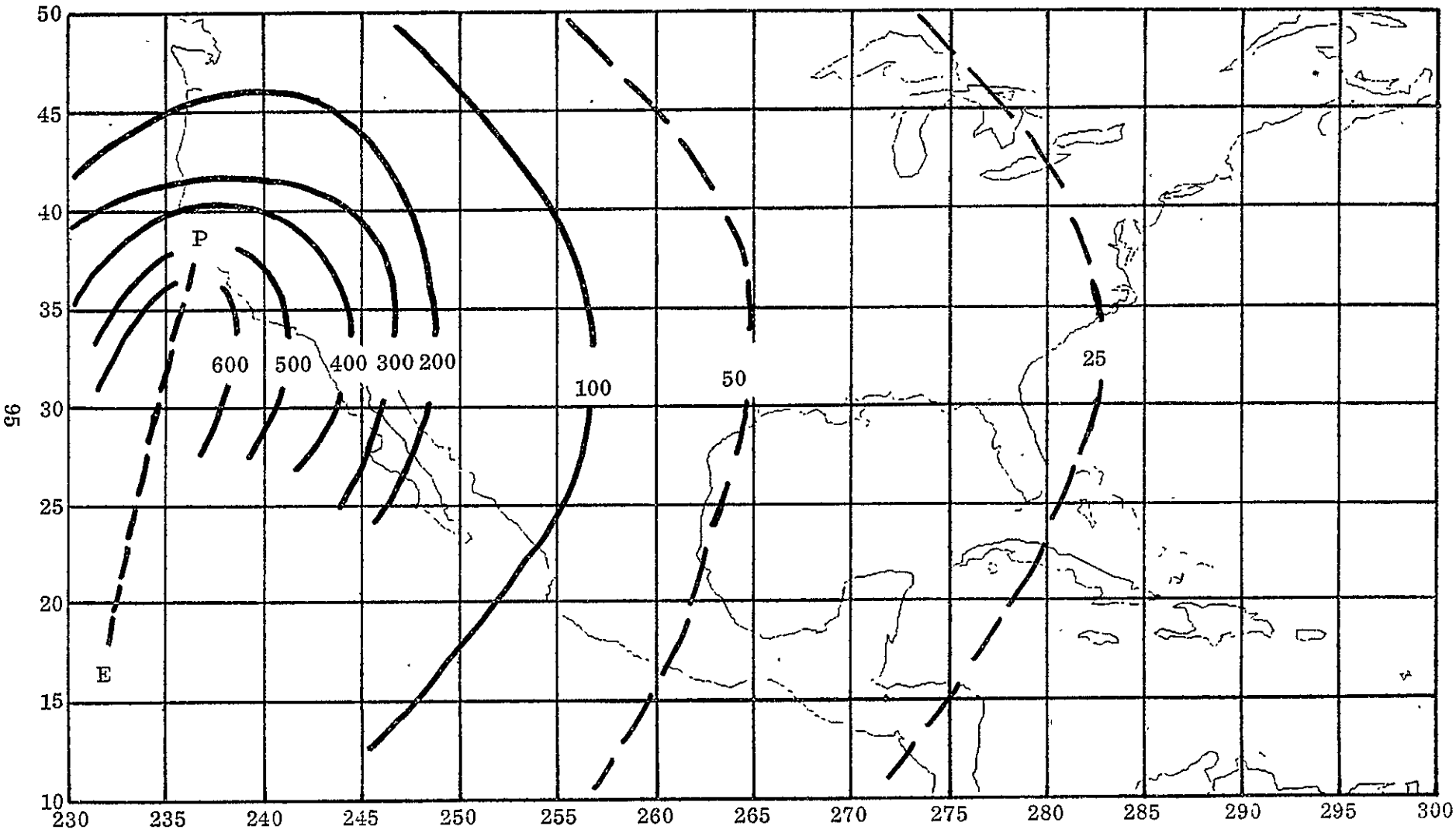

Fig. 5.10. Sensitivities of $\dot{\rho}$ in thousandths of a millimeter per second to φ in milligals. One satellite in a 700 km high orbit is tracked by a very high geostationary satellite. E denotes the position of the low satellite at the initial epoch, and P denotes its position at the epoch for which the partial derivatives are evaluated.

thus have little net effect on the range rate between them. While a single low satellite tracked by a high satellite approaches a block of positive density along the ground path of the orbit, the surface layer in that block continually attracts the satellite, thus increasing the velocity toward the block. After the satellite passes, it is pulled back and its velocity tends to decrease. However, the satellite has also been pulled downward into a lower orbit during the entire pass, which serves to increase its velocity. The net effect is an accumulative increase in velocity which is steepest during the time the satellite approaches the block and levels off as the satellite passes the block. However, this means that the blocks that have the greatest effect on the range rate are those far back on the ground path of the orbit, not those in the vicinity of the satellite position. Furthermore, all blocks very far back on the ground path will have approximately the same large effect on the range rate. This means that two satellites in this configuration cannot be expected to separate the values of the density in neighboring blocks as efficiently as the two satellites in the same low orbit. On the other hand, the densities in neighboring blocks can be separated by using orbits of different inclinations, or a combination of ascending and descending passes. Furthermore, the larger values of the sensitivity coefficients means that the uncertainties of the recovered values for the density of the surface layer should be smaller when this configuration is used.

To test the effect of the altitude of the satellite partial derivatives were evaluated for two satellites separated by 200 km in the same orbit 300 km high. In this case, the pattern of the sensitivity coefficients was the same as when the satellites were 700 km high. However, the magnitude of the two maxima along the ground path of the orbit was two to three times as large. This suggests that the sensitivities are approximately inversely proportional to the altitude of the two satellites, at least for this range of altitudes.

5.3 Experimental Solutions Using Simulated Data.

Several series of solutions were run to test the effects of the configuration of the satellites, the relative weighting of the different equations, the rate at which observations are taken, and the density of observational data. These solutions were designed to answer the questions raised in Chapter 1.

5.31 Experiments with 700 km High Orbits. The first series of solutions used the gravity field described by 92 $5^\circ \times 5^\circ$ blocks. Observations of range rate and position were generated at intervals of 32 seconds for nine passes of a pair of satellites separated by 200 km in the same orbit at an altitude of 700 km. All passes were arcs of a circular orbit with an inclination of 80°, and all were ascending passes. The first three solutions tested the effect of the relative weighting of the range rate and position observations.

Solution 1.1 used weights based on standard deviations of 0.05 mm/sec in range rate and 100 m in all components of position for both satellites. The uncertainties of the recovered values of the density parameters were quite large, ranging from two to 20 mgals. Since the largest value of the density parameter in any block was 7.2 mgals, this was judged an unsatisfactory solution.

For Solution 1.2, the standard deviation of the position observations was decreased to 10 m in each component. This significantly decreased the standard deviations of the recovered parameters, the largest uncertainty in any block being 4.6 mgals. The correlations between neighboring blocks was smaller and the conditioning of the normal equation matrix was better than in the previous solution. This showed that satellite position observations significantly add to the solution if the tracking accuracy is sufficient to determine satellite position to 10 meters. However, the numerical errors in the recovered parameters were excessively large, reaching 21 mgals in one block.

This solution appeared to indicate that position observations with an

uncertainty of 10 meters contain about the same information as satellite to satellite range rate observations with an uncertainty of 0.05 mm/sec. Were this so, the use of satellite to satellite tracking would be unnecessary, since future ground based tracking equipment utilizing pulsed laser systems will be able to determine satellite position to much better than 10 m. Therefore, a third solution was run utilizing only the position observation equations. This solution was overcome with numerical error, so that all significant digits were lost. This means that the matrix of coefficients of the reduced normal equations was so poorly conditioned that a solution was not possible with the algorithm used, which indicates that the position observations alone were not capable of resolving the gravity field in $5^\circ \times 5^\circ$ blocks.

One cause of the large uncertainties and correlations in the first two solutions appeared to be a lack of sufficient data. The number of range rate observations was only slightly greater than the total number of gravity field and orbit unknowns. Furthermore, the spacing between the ground tracks of the orbits was about twice the width of the blocks. Therefore, five more ascending passes and seven descending passes were added to the data set. Each block was then traversed by the ground path of at least one pass. The pattern of these passes is shown in Figure 5.11. With a data rate of one observation every 32 seconds, at least two range rate observations were made over each $5^\circ \times 5^\circ$ block.

Solution 1.4 utilized this larger data base. The weights were based on observational uncertainties of 0.05 mm/sec in range rate and 100 m in position. This solution was a significant improvement over those obtained with the smaller data set. The uncertainties of the recovered values of the density are shown in Figure 5.12, and the numerical errors in the solution are shown in Figure 5.13. Comparison with the "true" values of the density in Figure 5.4 shows that these uncertainties are still much larger than the values of the parameters to be recovered in many of the

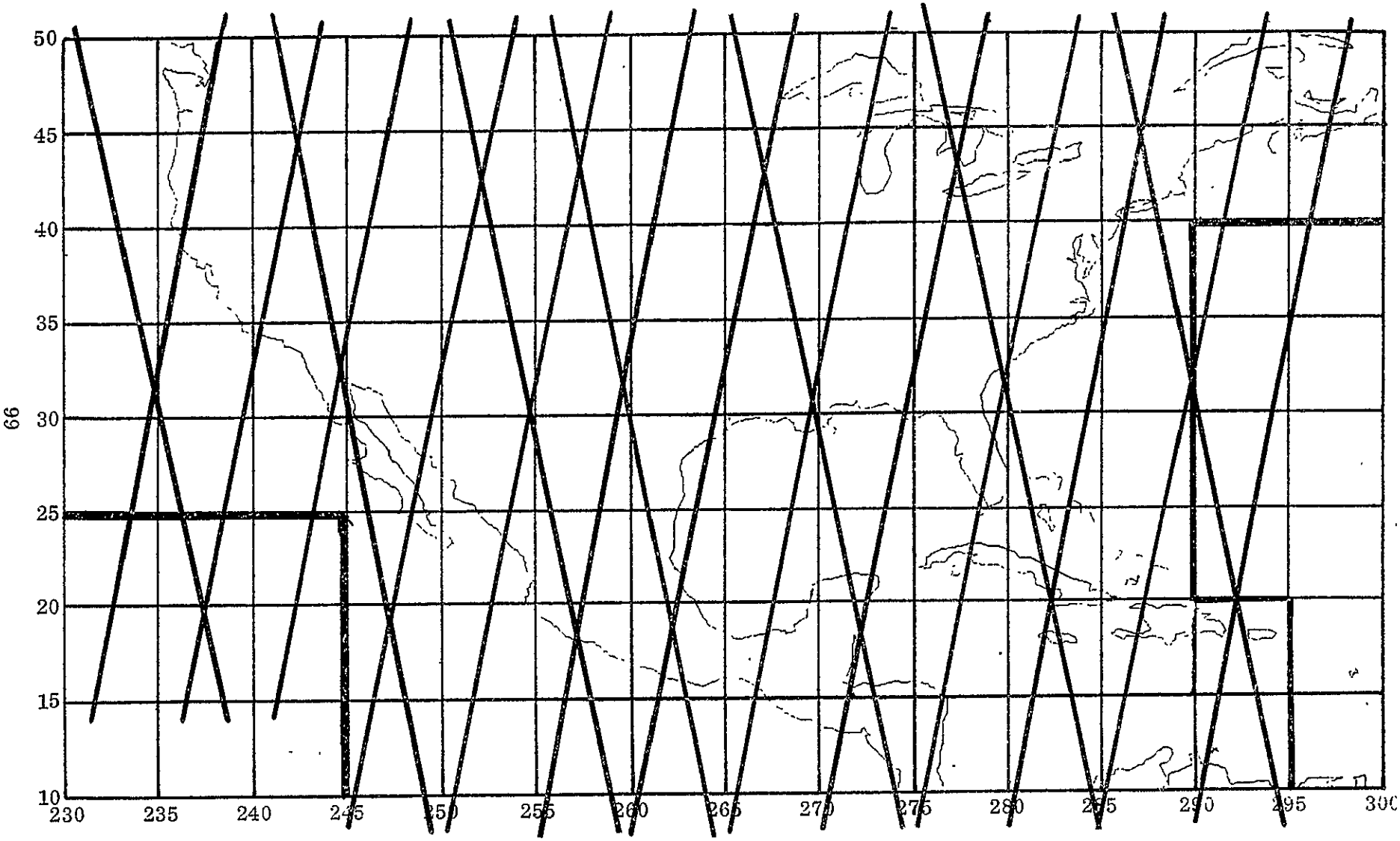

Fig. 5.11. Pattern of passes at 700 km altitude over the set of 5° x 5° blocks.

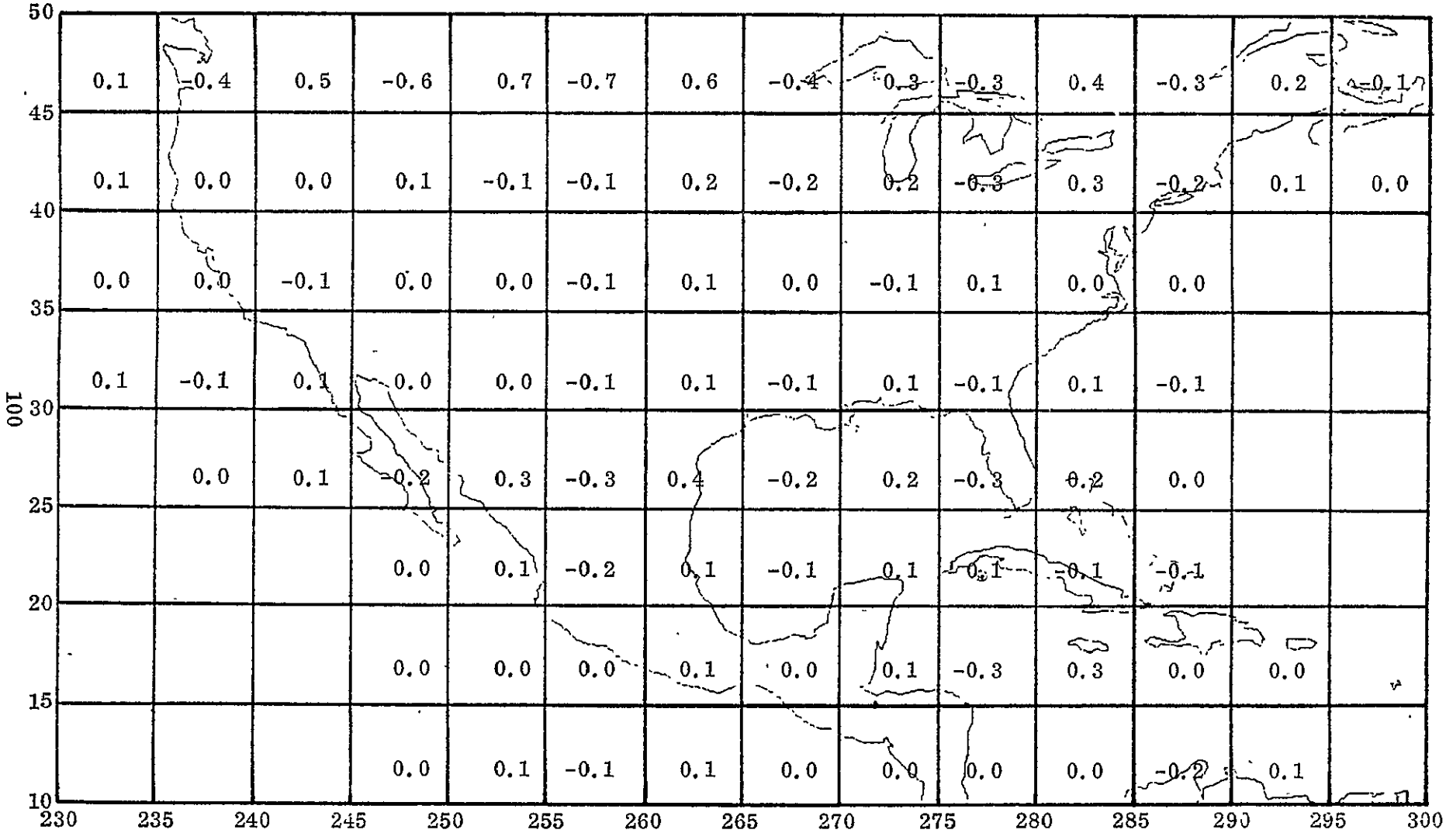

Fig. 5.13. Solution 1.4. Numerical errors in recovering density parameters in 5° x 5° blocks (mgals).

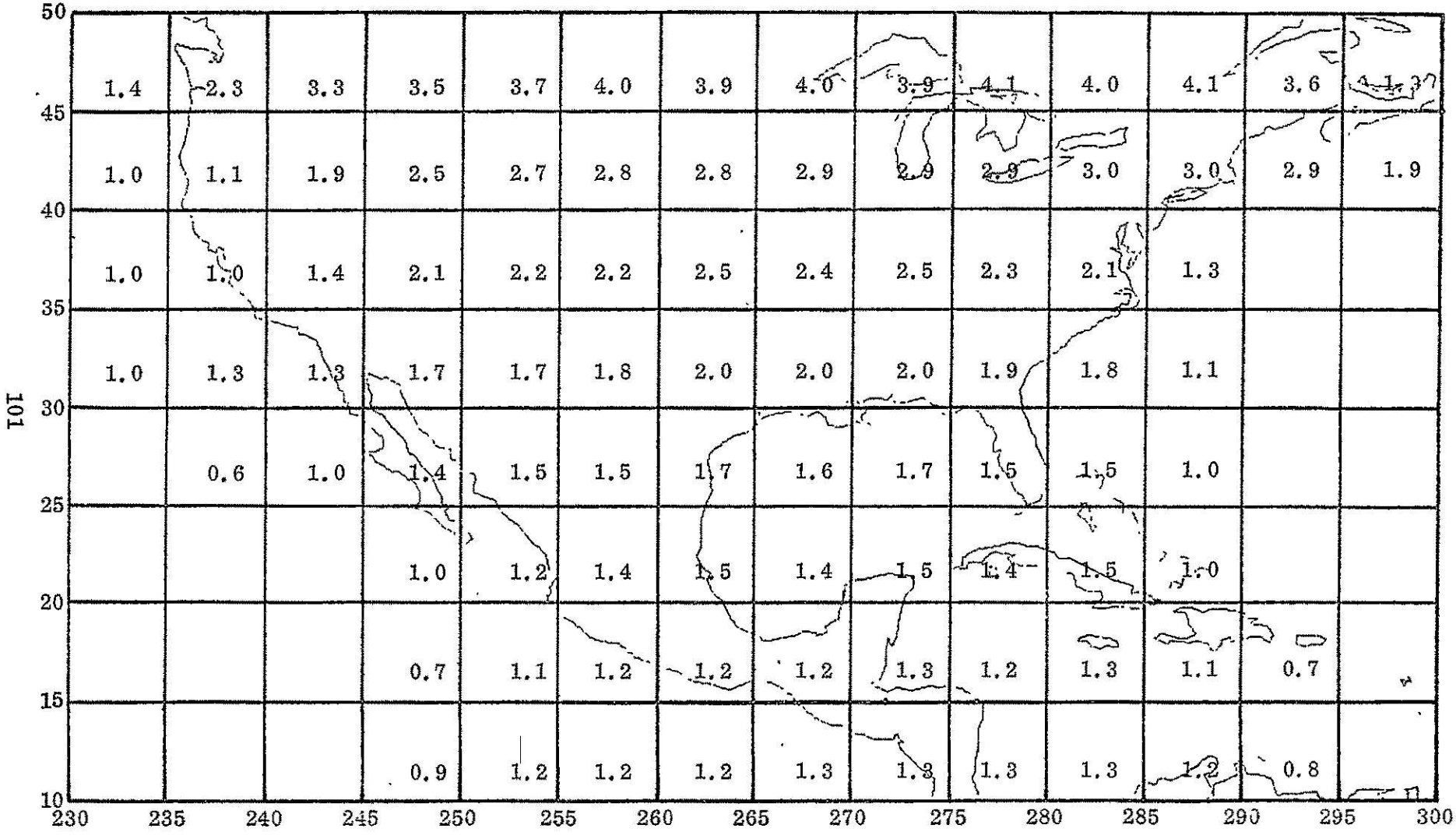

Fig. 5.12. Solution 1.4. Uncertainties of the recovered values of the density parameter in $5^\circ \times 5^\circ$ blocks (mgals).

blocks. The numerical errors are also larger than the "true" values in many cases, so that this solution cannot be judged completely satisfactory. However, the condition number of the coefficient matrix and the correlation coefficients indicate that the effects of neighboring blocks are successfully separated. The correlation coefficient between a block and its neighbor on the east or west ranged from -0.70 to -0.90. If one block intervenes between the two neighbors, the correlation is about +0.50. The correlation between a block and its neighbor to the north or south ranged from -0.20 to +0.25. The numerical errors in Figure 5.13 also indicate that the error in the recovered density in a block is significantly correlated with the error in a block to the east or west, but fairly independent of the error in a neighboring block to the north or south. Although the uncertainties of the recovered parameters are large, the successful separation of neighboring blocks indicates that a satisfactory solution could be obtained if more data were used.

Figures 5.12 and 5.13 also show a significant edge effect, with both the uncertainties and the numerical errors increasing toward the center and toward the north. The northward increase may be due to the fact that the passes extended somewhat farther to the south than to the north of the area under consideration. However, this same pattern was also evident in other solutions using different sets of passes. A more satisfactory explanation may be that the blocks in the north are somewhat smaller in area, so that their effects are somewhat harder to separate.

5.32. Experiments with 300 km High Orbits and Various Weights. A second set of simulated solutions was run to confirm and extend the results of the first series. In order to economize on the computer time required to run the simulations, a subset of 24 of the 92 $5^\circ \times 5^\circ$ blocks was used to describe the gravity field. Ten passes of a pair of satellites separated by 200 km were generated. Five ascending and five descending passes

completely covered the area under consideration. All passes were arcs of a circular orbit with an altitude of 300 km and an inclination of 80°. The lower altitude and denser coverage were expected to result in improved solutions.

Solution 2.1 confirmed this expectation. The weights for this adjustment were generated from standard deviations of 0.05 mm/sec for range rate observations and 100 meters for position observations. The uncertainties of the recovered parameters are shown in Figure 5.14 and the numerical errors are shown in Figure 5.15. The correlations between the recovered values of the density in neighboring blocks is described by the typical correlation pattern below.

1.0	-0.70	0.40
0.25	-0.22	0.11
0.30	-0.20	0.10

This pattern is interpreted by imagining the number in the upper left hand corner to be the correlation of a block with itself. The number in the first row and second column is the correlation of a block with its immediate neighbor to the east or west, and the number in the first row and third column is the correlation of a block with another block in the same latitude when a third block intervenes between them. Going down the columns describes correlations of a block with other blocks to the north or south in the same manner. These correlation coefficients are not computed for any particular block, but are typical of all the blocks. The actual correlation coefficients for any particular block may vary as much as ± 0.10 from these numbers.

The correlation coefficients for this solution show that neighboring blocks are fairly well separated, although there is still a significant negative correlation between blocks neighboring to the east or west. The pattern of the numerical errors in the solution also reflects the negative correlation

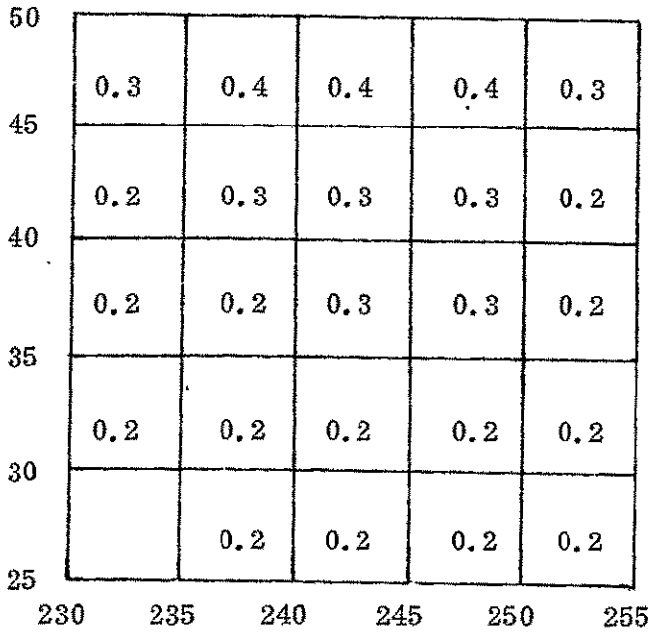

Fig. 5.14. Solution 2.1. Uncertainties of recovered parameters (mgals).

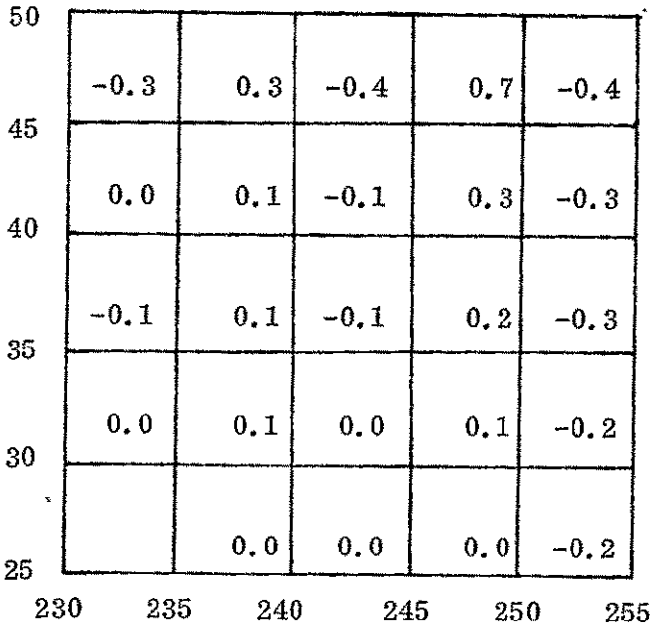

Fig. 5.15. Solution 2.1. Numerical errors in recovering density parameters (mgals).

between blocks immediately neighboring to the east or west and the positive correlation between blocks neighboring to the north or south. The uncertainties again show a definite edge effect, and an unexplained tendency to increase toward the north. The matrix of coefficients of the reduced normal equations was judged to be quite well conditioned, and altogether this solution was judged to be satisfactory.

The effect of different weightings of the observation equations was further tested. The standard deviation of the range rate equations was held at 0.05 mm/sec, and three solutions were run in which the standard deviations of the position observations were 10 m, 500 m, and 1000 m. The effect of this variation of the weights was extremely small. Between the extremes of 10 m and 1000 m for the standard deviations of the position observations, the uncertainties of the recovered parameters increased only by about 0.01 mgal and the numerical errors decreased by about the same amount. The condition of the normal equation matrix and the correlation coefficients showed almost no change. Another solution was run in which one satellite was assumed to be tracked with an accuracy of 100 meters and the other with an accuracy of 200 m. The results of this solution were practically identical with those of the solution in which both satellites were tracked with an accuracy of 100 meters. These solutions conclusively demonstrate that the accuracy of ground tracking of the two satellites has little effect on the solution. Precise tracking of the satellites from the ground would therefore be completely unnecessary.

Three more solutions tested extreme values of the weights. When a standard deviation of one meter was assigned to the position observations, the solution was overcome with numerical error and meaningless numbers were obtained. This also occurred when a weight of zero was assigned to the range rate observations and a standard deviation of ten meters was assigned to the position observations. These two experiments show that the position observations alone are not capable of resolving the components

of the gravity field, and if they are given too high a weight they may numerically overwhelm the range rate observations in the computer and ruin the solution. Therefore, precise tracking of the satellites from the ground is not only unnecessary, but also undesirable.

Finally, a solution was run in which the position observations were not used. This solution also broke down, showing that the position observations are necessary. In conclusion, it appears that the gravimetric information is contained in the range rate observations; tracking of the satellite positions is necessary to assign a geographic position to the phenomenon being observed, but contributes little else to the solution.

5.33. Experiments with 300 km Orbits and Various Configurations of the Two Satellites. Another series of simulations was designed to investigate the effect of varying the relative positions of the two satellites. Orbits at an altitude of 300 km were again used, but the configuration of the two satellites was varied. First the distance between two satellites in the same orbit was varied and the sensitivities of the range rate to the density of the surface layer in the 24 $5^\circ \times 5^\circ$ blocks were examined. When the satellites are brought closer together than 200 km, the magnitude of the maximum sensitivities decreases. This is because even blocks near to the subpoints of the satellites affect both satellites in approximately the same way, with no net effect on the range rate. When the distance between the two satellites was increased to 600 km, the magnitude of the maximum sensitivities also increased. However, the pattern also began to widen out somewhat, so that the separation of the effects of neighboring blocks was not quite as sharp.

Observation equations were also generated for the case when the satellites are close together in the same orbit plane, but one orbit is slightly higher than the other. The range rate between two satellites in this configuration was also found to be sensitive to the density of the surface layer. However, the observed range rate is primarily in the along track

direction rather than the radial direction even in this case, since the lower satellite tends to catch up with and pass the higher satellite during the pass. Thus, the observation equations generated by two satellites in this configuration are similar to those generated by two satellites in the same orbit.

Finally, observation equations were generated for two satellites in different orbital planes passing over the area under consideration at an altitude of 300 km and roughly 500 km apart. Because of the convergence of the two orbits, the actual distance between the two satellites varied from almost 600 km to near 300 km during a pass. The rate of change of the distance between the two satellites is also quite sensitive to the perturbing influence of the density layer in this case. Furthermore, the pattern of the sensitivity coefficients is quite different from the case of two satellites in the same orbit.

It appeared that the use of two satellites in this side by side configuration might help separate the effects of blocks neighboring to the east or west. Three passes of the two satellites in this configuration were generated and added to the data set used in Solution 2.1. The adjustment of these observations showed that the new passes did indeed help to separate the effects of neighboring blocks. The uncertainties of the adjusted density parameters were all under 0.2 mgal, and the matrix of normal equations was very well conditioned. The correlation of a block with its immediate neighbor to the east or west ranged from -0.14 to -0.76, with most of these correlation coefficients clustering around -0.45. All other correlation coefficients were below 0.30 in absolute value. Thus, a considerable improvement was obtained by adding only three new passes to the data set used in Solution 2.1.

Several conclusions may be drawn from the results of this set of experiments. First, it is not at all necessary that both satellites be in exactly the same orbit; some variation in their relative configuration is not only permissible but (highly) desirable. However, the distance between the two

satellites should not be too small nor too large. A range of 1/2 to 2 times the size of the blocks to be resolved appears to be a reasonable guideline.

Second, it is not necessary to have orbits of different inclinations. Thus, a single pair of satellites in low orbits at high inclinations could survey the gravity field of the entire earth.

Third, even making some allowance for edge effects, it appears that the Doppler measurement between two satellites in 300 km orbits can very effectively separate the gravity field in $5^\circ \times 5^\circ$ blocks. Furthermore, the mean values of the density parameter in these blocks can be determined to accuracies of better than 0.2 mgals, which corresponds to an accuracy of about one milligal in mean gravity anomalies.

In order to be sure that these conclusions would still hold for a larger area, a set of 18 passes of the two low satellites over the 92 $5^\circ \times 5^\circ$ blocks was generated. All orbits were 300 km high with an inclination of 80°. Seven ascending passes and five descending passes had the configuration of one satellite behind the other by 200 to 600 km. For six ascending passes the two satellites were in different orbital planes, traveling side by side on slowly converging orbital paths between 300 and 600 km apart. Range rate and position observations were again simulated every 32 seconds, so that at least two observations were taken over each block. These observations were adjusted with a standard deviation of 0.05 mm/sec assigned to the range rate observations. The uncertainties of the recovered density parameters in $5^\circ \times 5^\circ$ blocks are shown in Figure 5.16. These uncertainties are as low as those obtained in Solution 2.1 (Figure 5.14), and, in most cases, are significantly smaller than the parameter to which they apply (Figure 5.4). The numerical errors in recovering the density parameters were all below 0.2 mgal in absolute value, and most were below 0.1 mgal. The correlation between neighboring blocks varied considerably, but the pattern shown below is fairly representative.

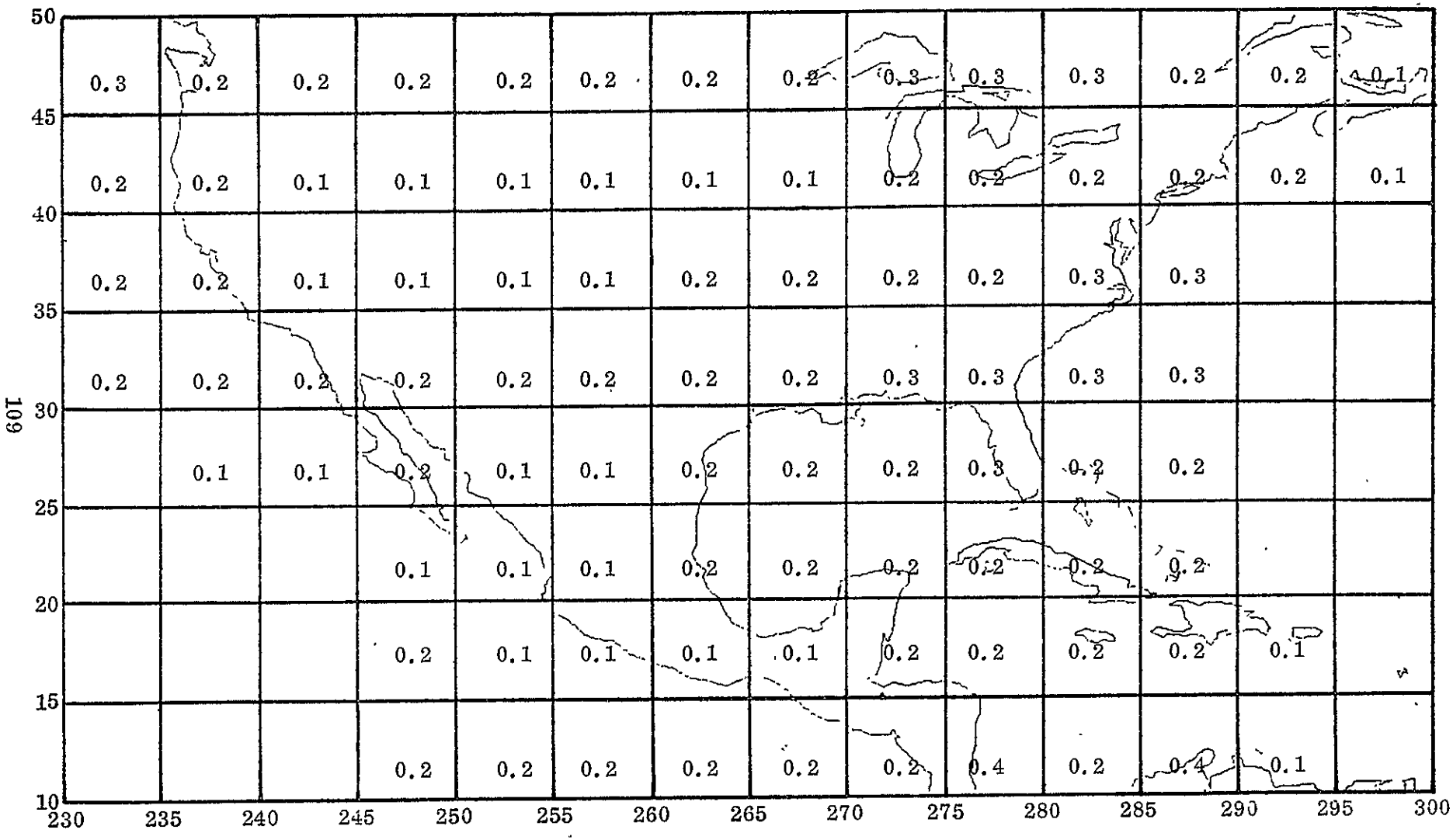

Fig. 5.16. Solution 3.1. Uncertainties of the recovered values of the density parameter in $5^{\circ} \times 5^{\circ}$ blocks (mgals).

1.0	-0.60	+0.20
+0.40	-0.30	+0.20
+0.40	-0.30	+0.15

The correlation coefficients between a block and its immediate neighbor to the east or west varied from -0.03 to -0.86; however, most of these correlation coefficients were between -0.30 and -0.70, indicating a good separation of neighboring blocks. The normal equation matrix was well conditioned, and the solution was judged to be satisfactory. This confirms the conclusion that doppler measurements between two satellites in 300 km high orbits can effectively resolve the gravity field in $5^\circ \times 5^\circ$ blocks.

5.34. Resolution of $2^\circ \times 2^\circ$ Blocks with Orbits 200 km High. The next series of experiments utilized the gravity field described by the mean values of the density parameter in $2^\circ \times 2^\circ$ blocks (Figure 5.8). A series of orbits passing over the area under consideration at an altitude of 200 km was generated. Each of the orbits was again circular with an inclination of 80°. This set of orbits included both ascending and descending passes, with one satellite behind the other by 200 km in some cases and to the side by about 200 km in others. Observations of range rate were generated every 32 seconds.

Since observations of the positions of the satellites do not contribute significantly to the solution for the parameters describing the gravity field, observations of this type were formed only every 128 seconds. The standard deviation assigned to the range rate observations was 0.05 mm/sec, and that assigned to the position observations was 100 m. This data was used to solve for the values of the density parameter in $2^\circ \times 2^\circ$ blocks, and the solution was designated Solution 5.1. The uncertainties of the recovered parameters are shown in Figure 5.17 and the numerical errors are shown in Figure 5.18. The "true" values of these parameters are shown in Figure 5.8. The uncertainties of this solution are about 10 times larger

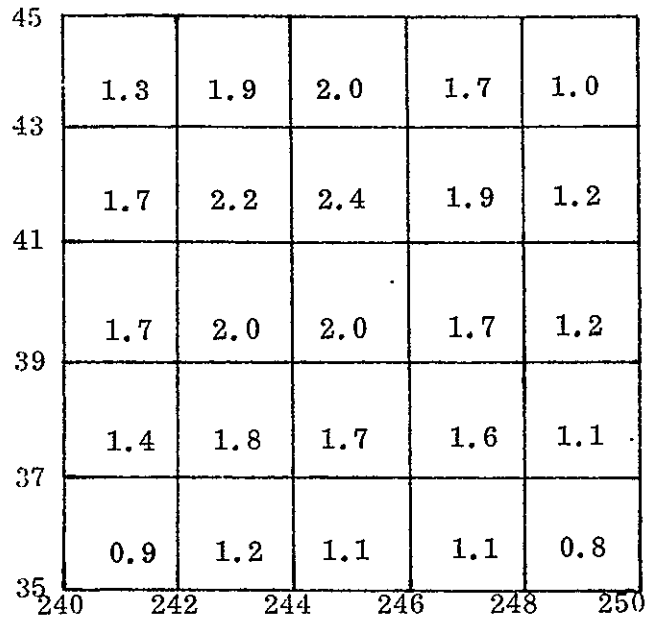

Fig. 5.17. Solution 5.1. Uncertainties of the recovered values of the density parameters in $2^\circ \times 2^\circ$ blocks (mgals).

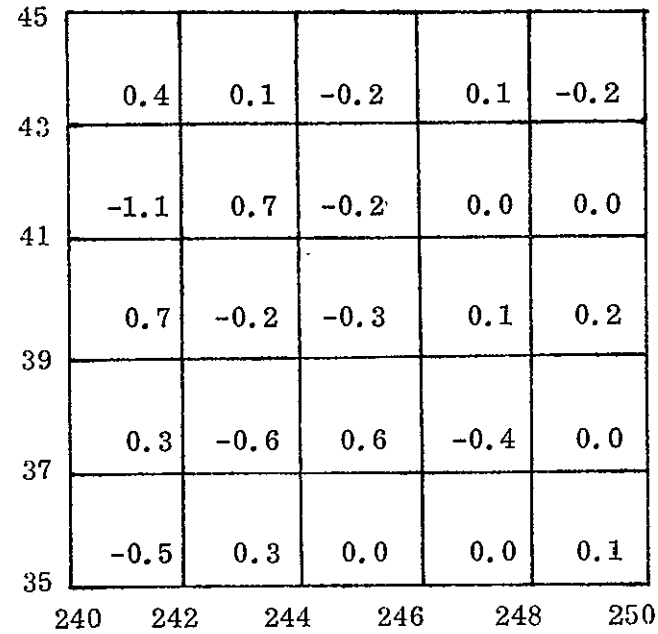

Fig. 5.18. Solution 5.1. Numerical errors in the recovery of the density parameters in $2^\circ \times 2^\circ$ blocks (mgals).

than those obtained when the gravity field in $5^\circ \times 5^\circ$ blocks was resolved from a satellite altitude of 300 km. This appears to indicate that the range rate observations cannot resolve the gravity field parameters in $2^\circ \times 2^\circ$ blocks from this altitude. However, the other statistics indicated a quite satisfactory adjustment. The condition of the normal equation matrix was quite good, and the correlation coefficients indicated fairly good separation of neighboring blocks. A typical pattern is shown below.

1.0	-0.80	+0.45
-0.60	+0.40	-0.20
+0.35	-0.20	+0.10

There was little variation among the correlation coefficients, and the actual correlation coefficients for a particular block were within 0.1 of the number shown in the typical pattern in all cases. Since neighboring blocks appeared to be reasonably well separated, this solution was judged to be marginally satisfactory. It was assumed that the uncertainties could be improved by using more data.

Additional data can improve an adjustment in two ways. Observations which contain new geometrical information help to separate the unknown parameters. Observations containing the same geometrical information as the old observations will contribute only statistical information; they will reduce the uncertainties of the recovered parameters but not the correlation coefficients.

For a satellite in a circular orbit at 200 km altitude, a data rate of one observation every 32 seconds corresponds to a rate of about one observation for every 2° of arc. This means that only a single observation was formed over many of the blocks in this data set. To test the effect of additional observations, the set of ten orbits was integrated again, but this time range rate observations were formed every 16 seconds, resulting in a set containing twice as many observations and at least two observations

over each of the $2^\circ \times 2^\circ$ blocks. The adjustment of this set of data was designated Solution 7.1. The uncertainties for this solution are shown in Figure 5.19 and the numerical errors are shown in Figure 5.20. The uncertainties are very slightly less than $1/\sqrt{2}$ times the uncertainties obtained in Solution 5.1 (Figure 5.17). An improvement by this factor would be expected even if the original observations were merely repeated without adding any new geometrical information to the data set. Furthermore, the condition number of the normal equation matrix is only slightly better in Solution 7.1 than in Solution 5.1. Similarly, the added observations lead to only a very slight improvement in the correlation coefficients. Thus, it appears that the higher data rate of one observation every 16 seconds does not significantly help separate the effects of neighboring $2^\circ \times 2^\circ$ blocks. For the same reason, higher data rates will not make possible the resolution of smaller blocks.

A simulation was performed in this series to test the way the algorithm means the effect of the disturbing gravity field into $2^\circ \times 2^\circ$ blocks. The assumed values of the parameters in the 25 $2^\circ \times 2^\circ$ blocks were the means of corresponding values in the set of 100 $1^\circ \times 1^\circ$ blocks which covered the same area (Figure 5.6). The smaller blocks more accurately describe the gravity field, since they can represent smaller features. Thus, orbits integrated using the disturbing potential based on the 100 $1^\circ \times 1^\circ$ blocks more accurately represent the actual motion to be expected of real satellites. For this experiment, data was generated from orbits integrated with the $1^\circ \times 1^\circ$ blocks, but the unknowns of the adjustment were the values of the density parameters in the $2^\circ \times 2^\circ$ blocks. If the algorithm meaned the effects of the disturbing gravity field properly, then the recovered values should be the same as when the $2^\circ \times 2^\circ$ blocks were used to generate the simulated data.

The ten orbits used in Solution 5.1 were integrated again, this time with the gravity field represented by $1^\circ \times 1^\circ$ blocks. Simulated observations

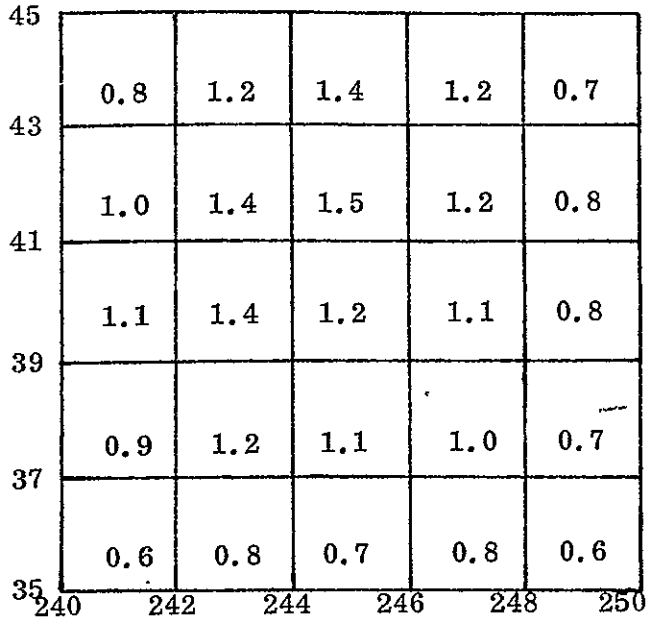

Fig. 5.19. Solution 7.1. Uncertainties of recovered values of the density parameters in $2^\circ \times 2^\circ$ blocks (mgals).

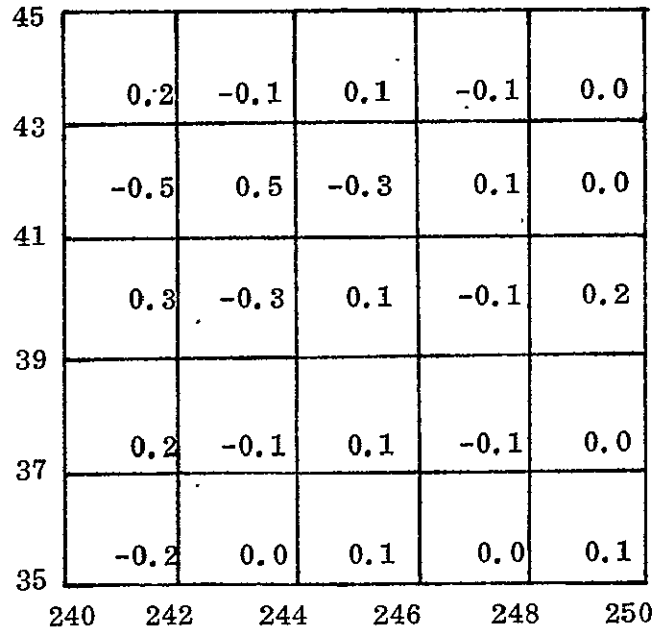

Fig. 5.20. Solution 7.1. Numerical errors in the recovery of the density parameters in $2^\circ \times 2^\circ$ blocks (mgals).

of range rate were again formed every 32 seconds. Observation equations, in which the gravity unknowns were the values of the density parameters in the $2^\circ \times 2^\circ$ blocks, were then formed. The adjustment of this data was designated Solution 6.1. Figure 5.21 shows that the uncertainties in this case are identical to those obtained in Solution 5.1 (Figure 5.17). Since the coefficients of the observation equations are the same in both cases, the correlation coefficients are also identical. Figure 5.22 shows the difference between values of the parameters in $2^\circ \times 2^\circ$ squares obtained from this solution and those of the "true" gravity field obtained by meaning the values in $1^\circ \times 1^\circ$ blocks. These errors are significantly larger than those in Solution 5.1 (Figure 5.18), and in many cases the numerical error is larger than the uncertainty associated with the same block. This demonstrates that the way the algorithm means the effects of the disturbing gravity field is indeed a significant source of error. In any solution utilizing real data this will therefore be a serious cause of concern.

5.35. Solutions Involving a Low Satellite Tracked by a High Geostationary Satellite. The discussion of the patterns of the sensitivity coefficients in Section 5.2 indicated that the configuration of a low satellite tracked by a geostationary satellite might not be able to resolve the gravity field as efficiently as two low satellites. To test this possibility numerically, ten orbits were integrated with the satellites in this configuration, using the gravity field described by the 25 $2^\circ \times 2^\circ$ blocks. The orbits for the low satellite were the same as those used for both satellites in simulating the data for Solution 5.1. Thus, the low satellite passed over the area under consideration at an altitude of 200 km in an orbit with an inclination of 80°. Both ascending and descending passes of the low satellite were used, and each block was traversed by the ground path of at least one pass of the low satellite. The geostationary satellite was placed in the equator at a longitude of 240°, in the meridian of the western boundary of the area being considered. Thus, the range rate was measured

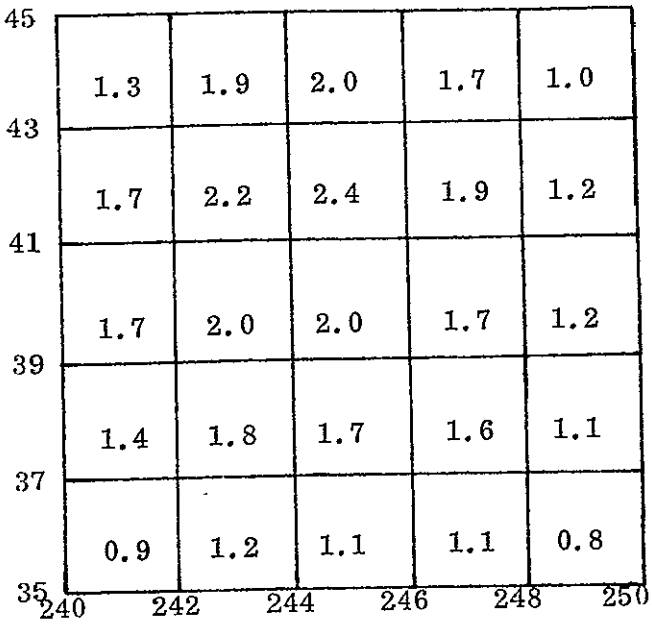

Fig. 5.21. Solution 6.1. Uncertainties of recovered values of the density paramters in $2^\circ \times 2^\circ$ blocks (mgals).

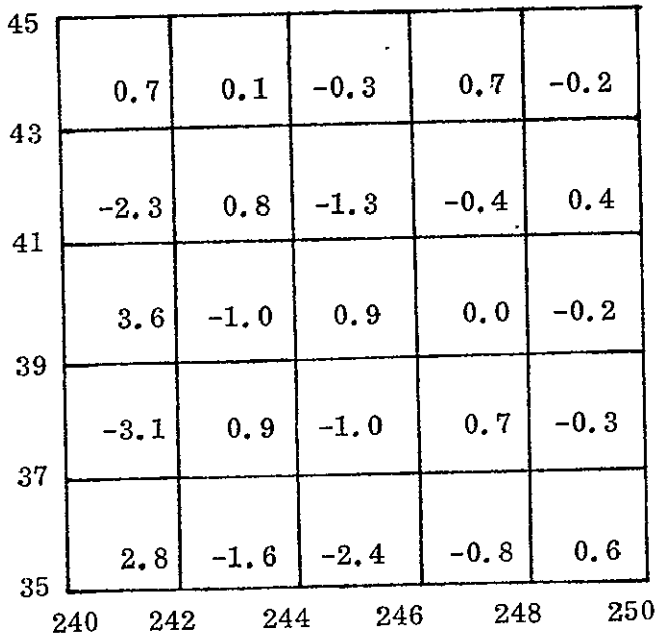

Fig. 5.22. Solution 6.1. Numerical error in the recovery of density parameters in $2^\circ \times 2^\circ$ blocks (mgals).

a direction inclined between 45° and 55° to the trajectory of the low satellite, so that both along track and radial perturbations were measured.

Range rate observations were generated every 16 seconds from these orbits, and position observations were generated every 128 seconds. Standard deviations of 0.05 mm/sec were assigned to the range rate observations and standard deviations of 100 m were assigned to the position observations. The adjustment of this set of data was designated Solution 8.1; the uncertainties and numerical errors are shown in Figures 5.23 and 5.24. The uncertainties of the recovered values of the density parameters in this case are only very slightly larger than those obtained in Solution 7.1 (Figure 5.19), which used the same amount of data but from two satellites in low orbits. A typical correlation pattern is shown below.

1.0	-0.80	+0.50	-0.35
-0.75	+0.60	-0.40	+0.25
+0.50	-0.40	+0.30	-0.15
-0.30	+0.25	-0.15	+0.10

The correlation pattern is extended to a four block by four block square, since the correlations approach insignificant values more slowly in this case. These correlation coefficients indicate that the separation of neighboring blocks is slightly worse than is the case with two satellites in low orbits. As further evidence of the slight worsening of the solution, the absolute magnitude of the largest correlation coefficient is slightly larger and the condition of the normal equation matrix is slightly worse than in the corresponding case with two low orbits. However, despite the slightly weaker solution, this experiment does indicate that satisfactory results can be obtained with the configuration of two satellites discussed in the Williamstown report [Kaula, 1969].

The concept of satellite to satellite tracking described in the Williamstown

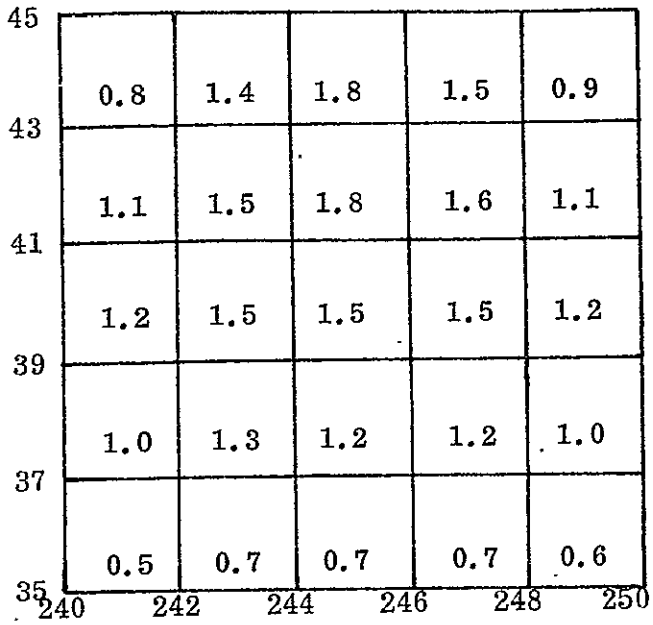

Fig. 5.23. Solution 8.1. Uncertainties of recovered values of the density parameters in $2^\circ \times 2^\circ$ blocks (mgals).

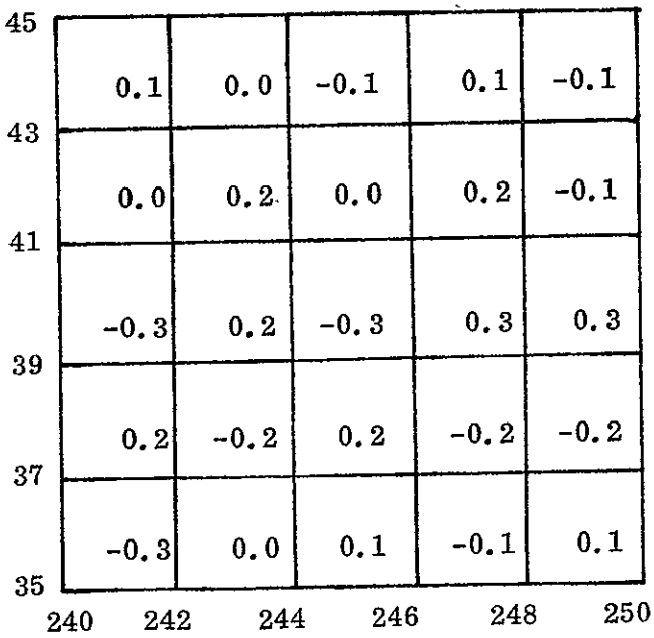

Fig. 5.24. Solution 8.1. Numerical errors in recovered values of the density parameters in $2^\circ \times 2^\circ$ blocks (mgals).

report includes three high satellites in geostationary orbits, placed along the equator 120° apart. With such a configuration, the low satellite will be visible to two of the geostationary satellites during much of its orbit. To test this concept, a second geostationary satellite was placed in the equator at 0° longitude, and simulated range rate observations were generated for the same ten passes of the low satellite used in the previous experiment. These observations were added to the set of observations from the first high satellite, resulting in a set of data twice as large as that used in Solution 8.1. It was expected that these new passes would contain significant new geometric information and would help to separate neighboring blocks in the same way as two low satellites side by side in slowly converging orbital paths. However, this was not the case. The uncertainties were smaller than those in Solution 8.1 (Figure 5.23) by a factor of $1/\sqrt{2}$, since the data set was twice as large; however, the improvements in the magnitudes of the correlation coefficients and in the condition of the reduced normal equation matrix were very slight. This indicates that a single high satellite, tracking twice as many passes, could resolve the gravity field as well as two high satellites.

The effect of varying the weights of the different observations was also tested for the configuration of a single low satellite tracked by high geostationary satellites. The assumed accuracy with which the lower satellite is tracked was changed to 200 m, and the accuracy with which the geostationary satellite is tracked was increased to 20 m. This corresponds to the situation in which the positions of the geostationary satellites are constantly monitored by ground based Very Long Baseline Interferometry and laser equipment while the orbits of the low satellite are only lightly tracked with radar. No significant change in the solution or any of its statistics were found when this weighting scheme was used. This, again, demonstrates that the gravimetric information is contained in the range rate measurement, and the position observations contribute little to the

solution for gravity parameters. However, if the geostationary satellites are precisely tracked for other purposes, this precise tracking will not overwhelm the range rate information and degrade the solution for the gravity parameters. Although this did occur in the case of two low satellites, the positions of the high geostationary satellites are almost completely insensitive to the perturbing effects of the surface layer. Thus, precise knowledge of the positions of high satellites cannot affect the solution for gravity parameters as can precise knowledge of the positions of low satellites.

5.36 Attempts to Resolve the Gravity Field in $1^\circ \times 1^\circ$ Blocks. All the previous numerical experiments indicated that the size of the smallest block that can be resolved depends almost linearly on the altitude of the low satellite. In Solution 1.4 (Figure 5.12), $5^\circ \times 5^\circ$ blocks were resolved from orbits 700 km high. Although the uncertainties of the recovered values of the density parameters were fairly high, the successful separation of neighboring blocks indicated that a satisfactory solution could be obtained by adding more of the same kind of data. In Solution 5.1 (Figure 5.17) $2^\circ \times 2^\circ$ blocks were successfully resolved from orbits 200 km high.

Attempts were made to resolve the gravity field in $1^\circ \times 1^\circ$ squares from the 200 km altitude, using both the low-low and the high-low configuration of the two satellites. The uncertainties of the density parameters recovered in these solutions are quite large, on the order of several hundred milligals. The normal equation matrices are also quite poorly conditioned, indicating weak solutions. These solutions were judged to be unsatisfactory.

In an attempt to find the altitude from which $1^\circ \times 1^\circ$ blocks can be resolved, a series of passes was integrated with the low satellite at the unrealistically low altitude of 100 km. The solution obtained with the low satellite at this altitude was still quite weak. The uncertainties of the recovered density parameters are on the order of 10 milligals. The

correlation coefficients are quite large, with many coefficients as large as 0.90 in absolute value. Furthermore, the correlation between two blocks falls off slowly with the distance between them, so that a recovered density in a block is still significantly correlated with that in another block which is four or five blocks away. These indicators show that the gravity field in $1^\circ \times 1^\circ$ squares cannot be successfully resolved even from 100 km altitude.

6. SUMMARY AND CONCLUSIONS

The experiments described in the previous chapter provided answers to the questions posed in the Introduction.

First, slightly better results may be obtained when a minimum altitude satellite is tracked by another satellite in nearly the same low orbit than when a cluster of geostationary satellites performs the tracking. On the other hand, solutions obtained by using the cluster of geostationary satellites are satisfactory, and this concept does offer several economic and operational advantages not afforded by the use of two low satellites. Since the high satellites are stationary, they can be tracked continuously. Furthermore, they can relay the measured Doppler count directly to the ground, thus obviating the need for data storage and delayed readout. The orbits of the high satellites are uneffected by air drag, so that only the low satellite need be equipped with a drag compensation device. The lifetimes of the geostationary satellites may be quite long; thus a single constellation of high satellites could conceivably be used to track several generations of minimum altitude satellites with short lifetimes. Finally, the configuration utilizing the geostationary satellites can perform many functions other than tracking the low satellites; the total concept of such a system is discussed in the Williamstown report [Kaula, 1969]. Because of these several advantages offered by this concept, the configuration of three geostationary satellites tracking a minimum altitude satellite is recommended.

If two satellites in low orbits are used, it is not at all necessary that efforts be made to maintain the circularity of the orbits or to maintain both satellites in precisely the same orbit. In fact, some variations in the relative configuration of the two satellites is desirable, since it adds significant geometric information to the solution. The only restriction on the relative configuration is that information is lost if the two satellites are too near together or too far apart. If geostationary satellites are

used to track the minimum altitude satellite, then variation in relative configuration of the satellites can be achieved if more than one high satellite is used. However, the use of a second high satellite did not appear to add significant geometrical information to the solution. Variation in the relative configuration of the satellites can also be achieved by using several low satellites in orbits of different inclinations. Although this possibility was not investigated, the patterns of the sensitivity coefficients suggest that the use of different inclinations would add significant geometric information and help to separate the gravity parameters in neighboring blocks. On the other hand, it is not necessary that several orbits of different inclinations be used, since the results obtained with a single low orbit are satisfactory.

The amount of gravimetric detail that can be resolved depends directly on the altitude of the low satellite. Even with a drag compensation device, the minimum altitude at which a satellite can remain in orbit for the length of time necessary to survey the gravity field on a global basis is about 200 km. From this altitude features in the gravity field as small as $2^\circ \times 2^\circ$ blocks may be resolved. This is true both when the minimum altitude satellite is tracked by another satellite in nearly the same low orbit and when a high geostationary satellite performs the tracking. With a density of data of about 2 observations for every block, the uncertainties of the recovered values of the density parameters in $5^\circ \times 5^\circ$ blocks were about one to three milligals when the low satellite was 700 km high. About the same range of uncertainties was obtained with about the same density of data when $2^\circ \times 2^\circ$ blocks were resolved from an altitude of 200 km. The $1^\circ \times 1^\circ$ blocks could not be satisfactorily resolved even from an altitude of 100 km. This suggests that the maximum altitude from which the effect of neighboring blocks can be separated is a function of the block size, but is not always exactly equal to the block size. A suggested curve giving the tradeoff between altitude and block size is shown in Figure 6.1.

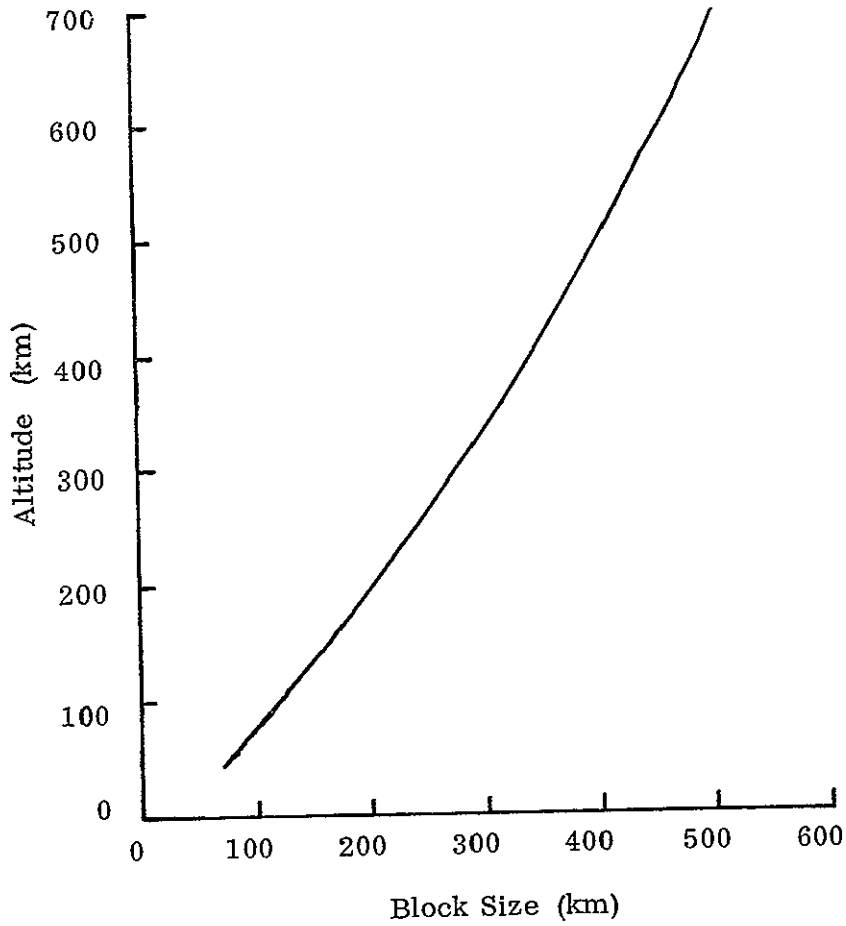

Fig. 6.1. Approximate Maximum Altitude From Which the Gravity Field Can Be Successfully Resolved, as a Function of Block Size.

Data rates of one observation every 32 seconds and one observation every 16 seconds were both successfully used to resolve the gravity field in $2^\circ \times 2^\circ$ blocks from 200 km orbits. This indicates that the 10 second averaging of the Doppler signal, necessary to achieve an accuracy of 0.05 mm/sec in range rate, will provide a rate of data sufficient for surveying the gravity field. In addition, a sufficient number of passes should be tracked so that the ground path of the low satellite traverses each of the blocks at least once.

The choice of a fictitious surface layer to represent the gravity field in these simulations was made purely for convenience. It is expected that the same results would have been obtained had gravity anomalies or gravity disturbances been used. Neither were the modifications of the conventional definitions of the normal and disturbing potential necessary. Comparison of Figures 5.3 and 5.4 shows that removal of the (12, 12) portion of the gravity field reduces the density of the fictitious surface layer only slightly. Therefore, the inclusion of the (12, 12) portion of the gravity field in the modified normal potential offers no real practical advantages. The conventional normal potential of a level ellipsoid could have been used just as effectively in these experiments, and the conventional definition would have allowed a much faster integration of the orbits and the partial derivatives. The experiments described here could have been performed just as effectively had a fictitious surface layer spread on the ellipsoid, rather than the surface of the earth, been specified. However, the use of the surface of the earth is more satisfying from a theoretical standpoint.

All of the simulations described involved short passes over a small portion of the earth's surface. The results that might be obtained from a global solution utilizing orbital arcs of several revolutions were not investigated. The most important factor affecting the resolution of the gravity field was the altitude of the low satellite, and it was shown that the possibility of resolving $2^\circ \times 2^\circ$ blocks depends on the possibility of maintaining a satellite in a 200 km high drag free orbit. It is entirely possible that the same resolution might be obtained if the position of a satellite in such a low orbit were precisely tracked from the ground over a very long arc of many revolutions. Although the study by Gaposchkin [1970] indicated that this resolution is not possible from the altitude of the present satellites which are equipped with laser retroreflectors, it may be possible from an altitude as low as 200 km.

Several incompletely solved problems remain in the algorithm discussed in the previous chapters. The most important is the effect of the unmodeled gravity field outside of the area under consideration. The neglect of the effect of the rest of the world outside of the area of interest will undoubtedly alias the solution for those parameters that are considered. However, routine computations with many thousands of unknowns are clearly impractical. This problem will not be completely solved until a local method of representation, more direct than the fictitious surface layer, is found. The theoretical work described in [Lundquist, et. al. 1970] offers some hope in this direction. In the meantime, it will be necessary to use localized areas for routine processing of satellite to satellite Doppler data. Solutions for the parameters describing the gravity field in localized areas can be used effectively for data screening. They can provide valuable and realistic information on the detailed structure of the gravity field in a local area of interest if they are constrained by independent knowledge of the low order components of the gravity field and by surface gravimetry. When carefully screened satellite to satellite range rate data providing global coverage becomes available, it will be reasonable to attempt a solution for a global description of the gravity field.

The effect of perturbations other than those caused by small features in earth's gravity field have not been considered at all here. Luni-solar perturbations are always present to some extent; although they are small, it may be necessary to account for their effects in the data reduction process. If two satellites in low orbits are used, the sun and moon will effect both in the same manner with no net effect on the range rate. However, a geostationary satellite revolves in an orbit a tenth of the way to the moon, so that it would certainly be affected by the moon differently than a minimum altitude satellite.

Even more critical is the extent to which a drag compensation device will be able to maintain a satellite in a purely gravitational orbit. In its

most elementary form, the air drag sensing system consists of a small unsupported proof mass contained in a small cavity at the mass center of the satellite. The external shell of the satellite shields the proof mass from air drag and radiation pressure, so that the proof mass follows a purely gravitational orbit. The sensing unit detects the motion of the rest of the satellite relative to the proof mass, and thrusters adjust the motion of the satellite to follow that of the reference proof mass. Deviations from a purely gravitational orbit occur only because the proof mass may be slightly perturbed by the sensor, by gradients in the local magnetic field, by mass attraction of the satellite, and by many other phenomena [Lange, DeBra, and Kaula, 1966]. The problem of most pertinence to the measurement of range rate is the possibility that the satellite will have some velocity relative to the proof mass. The satellite must move relative to the proof mass in order for the air drag to be sensed, and this relative motion must proceed at some velocity. Although this velocity will certainly be quite small, it could certainly reach a magnitude comparable to the projected noise level of 0.03-0.05 mm/sec in the range rate, and thus produce systematic errors in the range rate measurements. Therefore, the relative velocity of the proof mass and the main body of the satellite must be carefully considered in the design of the drag sensor and compensation device.

Another incompletely solved problem is the identification of the source of numerical error in the algorithm. Numerical errors were evident in all the simulated solutions, since the original values of the unknowns were never recovered to better than two or three significant digits. This was not considered a serious cause for concern, since the numerical errors were always smaller than the uncertainties associated with the same parameters. Several sources of numerical error may be suggested. Among these are (1) the roundoff and truncation errors involved in numerically integrating the differential equations for the orbits and the transition matrices,

as well as the error in numerically evaluating the definite integral for the parameter sensitivity matrix; (2) the numerical evaluation of integrals of the form $\int\int 1/\ell \, dS$; (3) the linearization of the observation equations; and (4) roundoff error accumulated in the solution of the normal equations. The fact that the observation equations were tested and found to be satisfactorily linear within the range of values expected suggests that the last error source may be a principle contibutor. This is not unexpected, since the solution of a large set of simultaneous linear equations almost always involves a considerable accumulation of roundoff error.

A closely related problem is the way in which the algorithm means the effect of the surface layer in a block. The experiment described in Section 5.34 showed this to be a serious cause of concern. This problem appears to be related to the numerical evaluation of the integrals of the form $\int\int 1/\ell \, dS$. When $1^\circ \times 1^\circ$ blocks were used to integrate the "true" orbit, the accelerations of the satellites were computed by summing the forces exerted by each of these blocks. However, the sum of four of these forces is not the same thing as the force exerted by a $2^\circ \times 2^\circ$ block whose density is the sum of the four densities in the $1^\circ \times 1^\circ$ blocks. The difference between the two forces leads to significant numerical errors in the solution. It appears that simple schemes for computing the mean effect of a block can only be used if the distance from the satellite to the block is much greater than the dimensions of the block, but in that case the observations will not be able to separate the effects of neighboring blocks. Therefore, it will be necessary to compute the mean effect of a block by schemes more sophisticated than evaluating functions at the midpoint of a block or even at the midpoints of four sub-blocks within each block.

The minimum altitude of 200 km implies that the use of satellite to satellite Doppler tracking cannot be expected to resolve blocks smaller than $2^\circ \times 2^\circ$, or features in the gravity field smaller than 200 km on a side.

This means that satellite to satellite tracking cannot replace surface gravity surveys. However, as a global surveying system, it can greatly increase our knowledge of the global structure of the gravity field. Global knowledge of the gravity field in $2^\circ \times 2^\circ$ equal area blocks will provide as much information as a spherical harmonic series complete through degree and order (90, 90). The accuracy with which the gravity field in $2^\circ \times 2^\circ$ blocks can be determined depends on the accuracy of the range rate measurement. With the projected tracking accuracy of 0.03-0.05 mm/sec, and with a sufficient number of observations, it should be possible to determine the parameter describing the fictitious surface layer to an accuracy of one milligal. In terms of mean gravity anomalies in $2^\circ \times 2^\circ$ blocks, this corresponds to an accuracy of about six milligals. This is considerably better than the accuracy with which the mean gravity anomaly in a $2^\circ \times 2^\circ$ block can be determined by measuring surface gravity along a single profile through the block, even when the least standard error method of interpolation of gravity anomalies is used [Moritz, 1963]. The accuracy with which the shape of the geoid can be determined from satellite to satellite Doppler data will depend to a large extent on the correlations between neighboring blocks as well as the accuracy with which the gravity field is determined in a single block. However, improved knowledge of the gravity field in $2^\circ \times 2^\circ$ blocks should help to achieve the goal of geoid determinations approaching the 10 cm accuracy required by oceanography. Thus, satellite to satellite Doppler tracking can considerably refine our knowledge of the gravity field, both by performing fairly detailed surveys of local ocean areas and by surveying the gravity field on a global basis. Because of the great potential of the method, development of both the instrumentation and the data reduction techniques should receive continuing attention.

References

Anderle, R. J., C. A. Malyevac and H. L. Green Jr. (1969). "Effect of Neglected Gravity Coefficients on Computed Satellite Orbits and Geodetic Parameters." Presented to the National Meeting of the American Geophysical Union, April. (Abstract in Transactions, American Geophysical Union, Vol. 50, No. 4).

Anderle, R. J. (1970). "Doppler GEOS-II." Presented to the GEOS-II Review Meeting. NASA-Goddard Space Flight Center, June 22-24.

Baker, Robert M. L. Jr. (1960). "Orbit Determination from Range and Range-Rate Data." Presented at the Semi-Annual Meeting of the American Rocket Society, Los Angeles, May 9-12.

Conte, Samuel D. (1962). "The Computation of Satellite Orbit Trajectories, in Advances in Computers, Vol. 3, (Frank L. Alt and Morris Rubinoff, editors). Academic Press, New York.

Conte, Samuel D. (1965). Elementary Numerical Analysis. McGraw-Hill, New York.

Danby, J. M. A. (1962). "Integration of the Equations of Planetary Motion in Rectangular Coordinates," Astronomical Journal, Vol. 67, p. 287.

Douglas, Bruce and James Marsh (1970). "Gravity Field Comparisons Based on GEOS-I and GEOS-II Orbital Analysis." Paper presented to the GEOS-II Review Meeting. NASA-Goddard Space Flight Center, June 22-24.

Faddeev, D. K. and V. N. Faddeeva (1963). Computational Methods of Linear Algebra. Translated by Robert C. Williams. W. H. Freeman, San Francisco.

Gaposchkin, E. M. (1970). "The 1969 Smithsonian Standard Earth and Global Tectonics." Presented to the GEOS-II Review Conference. NASA-Goddard Space Flight Center, June 22-24.

Gaposchkin, E. M. and K. Lambeck (1969). "New Geodetic Parameters for a Standard Earth." Presented to the 1969 National Fall Meeting of the American Geophysical Union, December. (Abstract in Transactions of the American Geophysical Union, Vol. 50, No. 11).

Gaposchkin, E. M. and K. Lambeck (1970). "1969 Smithsonian Standard Earth (II)." Smithsonian Astrophysical Observatory Special Report 315. Cambridge, Mass.

Hamming, R. W. (1959). "Stable Predictor Corrector Methods for Ordinary Differential Equations," Journal of the Association for Computing Machinery, Vol. 6, pp. 37-47.

Heiskanen, Weikko A. and Helmut Moritz (1967). Physical Geodesy. W. H. Freeman and Co., San Fransisco.

Heiskanen, W. A. and F. A. Vening-Meinesz (1958). The Earth and Its Gravity Field. McGraw-Hill, New York.

Hirvonen, R. A. and Helmut Moritz (1963). "Practical Evaluation of Gravity at High Altitudes." Reports of the Institute of Geodesy, Photogrammetry and Cartography No. 27. The Ohio State University, Columbus.

Hotine, M. and F. Morrison (1969). "First Integrals of the Equations of Satellite Motion," Bulletin Géodésique, No. 91.

Hudson, Edward F. (1970). "Determining the Geometric Shape of the Geoid in the Ocean Covered Areas of the World by Satellite Radar Altimetry." Presented to the 51st Annual Meeting of the American Geophysical Union, April 20-24, Washington D.C. (Abstract in Transactions, American Geophysical Union, Vol. 51, No. 4).

Kane, M. F. (1969). "Doppler Gravity, a New Method," Journal of Geophysical Research, Vol. 74, p. 6579.

Kaula, William M. (1966). "Tests and Combination of Satellite Determinations of the Gravity Field with Gravimetry," Journal of Geophysical Research, Vol. 71, p. 5303.

Kaula, William, W. H. K. Lee, P. T. Taylor, and H. S. Lee (1966). "Orbital Perturbations from Terrestrial Gravity Data." Institute of Geophysics and Planetary Physics, University of California, Los Angeles.

Kaula, William (editor), (1969). "Solid Earth and Ocean Physics." Report of a Study at Williamstown, Mass., to the National Aeronautics and Space Administration. Sponsored by NASA-Electronics Research Center, MIT-Measurement Systems Laboratory, Cambridge, Mass. August.

Koch, K. R. (1968). "Alternate Representation of the Earth's Gravity Field for Satellite Geodesy," Bollettino di Geofisica Teorica ed Applicata. Vol. X, No. 40.

Koch, Karl-Rudolf and Foster Morrison (1970). "A Simple Layer Model of the Geopotential from a Combination of Satellite and Gravity Data," Journal of Geophysical Research, Vol. 75, p. 1483.

Lambeck, Kurt (1970). "A Discussion of the 1969 Standard Earth." Presented to the 1970 Annual Meeting of the American Geophysical Union, Washington, D.C., April 22. (Abstract in Transactions, American Geophysical Union, Vol. 51, No. 4).

Lange, Benjamin (1964). "The Drag Free Satellite," AIAA Journal, Vol. 2, p. 1590.

Lange, Benjamin O. Daniel B. DeBra, William M. Kaula (1969). Final Technical Report on A Preliminary Design of a Drag-Free Satellite and its Application to Geodesy." Stanford University Center for Systems Research, Guidance and Control Laboratory, May.

Lundquist, C. A. and G. Veis (editors) (1966). "1966 Smithsonian Standard Earth," Smithsonian Astrophysical Observatory Special Report No. 200 (in 3 volumes). Cambridge, Mass.

Lundquist, C. A. (1967a). "Satellite Altimetry and Orbit Determination," Smithsonian Astrophysical Observatory Special Report 248. Cambridge, Mass.

Lundquist, C. A. (editor) (1967b). "Geodetic Satellite Results During 1967." Smithsonian Astrophysical Observatory Special Report 264. Cambridge, Mass.

Lundquist, C. A., G. E. O. Giacaglia, K. Hebb, and S. G. Mair (1969). "Possible Geopotential Improvement from Satellite Altimetry," Smithsonian Astrophysical Observatory Special Report 294. Cambridge, Mass.

Lundquist, C. A. (1970). "SAO Plans for the Use of GEOS-C for Earth Physics." Presented to the GEOS-II Review Meeting, NASA-Goddard Space Flight Center. June 22-24.

Moritz, Helmut (1963). "Accuracy of Mean Gravity Anomalies Obtained from Point and Profile Measurements," Reports of the Institute of Geodesy, Photogrammetry, and Cartography. Report No. 29. The Ohio State University. Columbus.

Morrison, Foster (1970). "Comments on Paper by Milo Wolff, 'Direct Measurements of the Earth's Gravitational Potential Using a Satellite Pair'," Journal of Geophysical Research, Vol. 75, No. 11.

Mueller, Ivan I. and John D. Rockie (1966). Gravimetric and Celestial Geodesy — A Glossary of Terms. Frederick Ungar. New York.

Muller, Paul M. and W. L. Sjogren (1968a). "Consistency of Lunar Orbiter Residuals with Trajectory and Local Gravity Effects," JPL Technical Report 32-1307.

Muller, P. M. and W. L. Sjogren (1968b). "Mascons: Lunar Mass Concentrations," Science, Vol. 61, p. 680. (Also reprinted as JPL Technical Report 32-1339).

Needham, Paul E. (1970). "The Formation and Evaluation of Detailed Geopotential Models Based on Point Masses." Dissertation. The Ohio State University.

Obenson, Gabriel F. T. (1970). "Direct Evaluation of the Earth's Gravity Field from Orbital Analysis of Artificial Earth Satellites," Reports of the Department of Geodetic Science No. 129. The Ohio State University. Columbus.

Pellinen, L. P. (1970). "Estimation and Application of Degree Variances of Gravity." Studia Geophyica et Geodaetica. Vol. 14, p. 168.

Ralston A. and H. Wolf (editors) (1960). Mathematical Methods for Digital Computers. Wiley and Sons, New York, pp. 95-109.

Rapp, Richard H. (1966). "The Equatorial Radius and the Zero Order Undulation of the Geoid," Reports of the Department of Geodetic Science No. 73. The Ohio State University. Columbus.

Rapp, Richard H. (1967). "Comparison of Satellite Geoids and Anomaly Fields." Reports of the Department of Geodetic Science No. 80. The Ohio State University. Columbus.

Rapp, Richard H. (1969). "The Geopotential to (14, 14) From a Combination of Satellite and Gravimetric Data," Bulletin Géodésique, No. 91.

Riley, James D., Morris M. Bennett, and Emily McCormick (1967). "Numerical Integration of Variational Equations," Mathematics of Computation, Vol. 21, p. 12.

Schwarz, Charles R. (1967). "Optimal Filter and Least Squares Estimation," Geodetic Memorandum No. 1610. U.S. Army Map Service, Washington, D.C.

Strange, W. E. and G. P. Woollard (1964). "The Prediction of Gravity in the United States Utilizing Geologic and Geophysical Parameters," Hawaii Institute of Geophysics Report 64-18. University of Hawaii.

Uotila, Urho A. (1967). "Introduction to Adjustment Computations with Matrices," Unpublished Notes. Department of Geodetic Science. The Ohio State University. Columbus.

Warner, Michael R. and Melba W. Nead (1965). "SPODP - Single Precision Orbit Determination Program," Jet Propulsion Laboratory Technical Memorandum No. 33-204. Pasadena.

Weiffenbach, G. C. (1967). "Tropospheric and Ionospheric Propagation Effects on Satellite Radio-Doppler Geodesy," in Electromagnetic Distance Measurement: A Symposium. University of Toronto Press.

Weiffenbach, George C. (1969). "The GEOS-C Radar Altimeter." Presented to the Second Marine Geodesy Symposium. Marine Technology Society. New Orleans. November 3-5.

Wong, Lim and R. Prislin (1970). "Comparison of Geopotential Models for GEOS-I Prediction." Presented to the 51st Annual Meeting of the American Geophysical Union, April. (Abstract in Transactions, American Geophysical Union, Vol. 51, No. 4).

Wolff, Milo (1969). "Direct Measurements of the Earth's Gravitational Potential Using a Satellite Pair," Journal of Geophysical Research, Vol. 74, p. 5295.