JPL Publication 98-15

An Analysis of Gravity-Field Estimation Based on Intersatellite Dual-1-Way Biased Ranging

J.B. Thomas
Jet Propulsion Laboratory, Pasadena, California

**National Aeronautics and
Space Administration**

**Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California**

May 1999

195pp

Report Documentation Page

Report Date 01MAY1999	**Report Type** N/A	**Dates Covered (from... to)** -

Title and Subtitle An Analysis of Gravity-Field Estimation Based on Intersatellite Dual-1-Way Biased Ranging	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) NASA Jet Propulsion Laboratory Pasadena, CA	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes	
Abstract	
Subject Terms	
Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU
Number of Pages 195	

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

ACKNOWLEDGMENTS

The author is particularly grateful to Ab Davis for his encouragement to persevere with the unconventional error analysis introduced in this report, when there was doubt about its value. Gratitude is also extended to Ab for supplying the initial version of Fig. 2-1. The author is grateful to Ron Muellerschoen for supplying the sequential-estimation software that was adapted in Section 6 to the hypothetical acceleration observable; to Srinivas Bettadpur for information and results from conventional gravity estimation; to Courtney Duncan for the analysis of J2 orbital effects; to Jeff Srinivasan for providing the initial version of Fig. 3-1, and to Tom Hamilton for many detailed comments including the suggestion that the dual-band ionosphere correction be applied to high-rate samples rather than filtered samples. Gratitude is also extended to George Purcell, Bob Treuhaft, Charley Dunn, John Ries, Larry Romans, Willy Bertiger, and Yoaz Bar-Sever for numerous helpful comments.

ABSTRACT

The GRACE (Gravity Recovery And Climate Experiment) mission is designed to make global, highly accurate measurements of the Earth's gravity field with high spatial resolution. Ancillary GPS occultation measurements are also to be carried out for atmospheric monitoring. In the dual-1-way biased ranging of this mission, the range between two satellites, separated by 100 to 200 km in nearly polar, coplanar, circular orbits, is measured to very high precision, to within an additive constant, through the exchange of K- and Ka-band sinusoidal signals. Such biased ranging data, along with GPS L-band range and phase data, can be processed and fit over successive multiday intervals to obtain accurate estimates of the Earth's gravity field. This report approximately models and analyzes this process, from the generation of the RF signals at the two satellites through the extraction of the geopotential. The steps include generation of the transmitted signals, processing the received signals to extract "high-rate" baseband phase, carrying out a dual-1-way combination of baseband phase to extract high-rate biased range for each band, combining K- and Ka-band ranges to correct for the ionosphere effect, and processing the resulting high-rate biased range values to extract three types of reduced-rate observables: biased range, range rate and range acceleration. The version of dual-1-way biased ranging developed by this report improves upon previous versions in a number of ways: highly accurate satellite-timetag corrections derived from concurrent GPS data, better baseband phase extraction using highly digital processing, highly accurate USO-rate calibration derived from concurrent GPS data, an improved method for extracting high-rate biased range from baseband phase, improved filtering for extracting reduced-rate observables from high-rate biased range, and parallel extraction of three observable types. Relative to the earlier versions of dual-1-way processing considered here, the observable-extraction filters proposed in this report lead to greatly reduced random errors and/or greatly reduced gravity-signal amplitude distortion. Spectral models for the random observable errors due to system noise and ultrastable oscillators are developed for range, range rate, and range acceleration. A crude hypothetical fitting approach based on the acceleration observable is developed to provide insight and to estimate the random gravity-field errors resulting from system noise, ultrastable oscillators and accelerometers. Based on the improved dual-1-way processing and revised instrumental performance, this hypothetical acceleration analysis suggested that per-degree geoid errors on the order of 10 μm (degree ≤ 50, 300-km altitude, 200-km range, 13-day fit) are feasible with respect to anticipated random errors due to system noise, USOs, and accelerometers. The large improvement in random geoid errors obtained by this acceleration-observable analysis made it possible, with respect to the indicated random errors, to propose measurement of the relatively small time-varying components of the gravity field, which greatly increased the potential value of the GRACE mission. The hypothetical acceleration approach is analyzed from the frequency-domain perspective to provide useful insights and results for gravity estimation. The frequency-domain approach is also applied to several relevant simple models to approximately assess the effectiveness and consequences of using a piecewise-polynomial function of time to remove $f^{-\alpha}$ instrumental noise, an approach that results in a number of useful insights and results concerning the selection of polynomial order and update-interval length, the "mapping" of tones in a partial for a given gravity coefficient to lobes in the corresponding frequency-domain fit filter for that coefficient, gravity-coefficient error amplification due to excessive fit-filter mainlobes and sidelobes, and the "separation" of orbit parameters from gravity coefficients in multiparameter fits. An analysis to assess the importance of correlations between parameters in the hypothetical acceleration fits suggests such correlations do not greatly increase per-degree geoid errors.

TABLE OF CONTENTS

U

LIST OF FIGURES

SECTION 2

SECTION 3

SECTION 4

SECTION 5

1

SECTION 6

APPENDIX A

8

APPENDIX B

APPENDIX C

9

APPENDIX D

APPENDIX E

11

SECTION 1

INTRODUCTION

The GRACE (Gravity Recovery And Climate Experiment) mission is designed to make global, highly accurate measurements of the Earth's gravity field with high spatial resolution. Ancillary GPS occultation measurements are also to be carried out for atmospheric monitoring. The primary goal is to measure the Earth's gravity field over a period of approximately 5 years with sufficient accuracy (better than 0.1 mm geoid, per degree), frequency (once every 13 to 26 days), and spatial resolution (100 to 200 km, half wavelength) to provide valuable information concerning not only the static field but also its variable components due, for example, to mass-distribution changes in the ocean, ground water, and solid earth. Two satellites in nearly coplanar, circular, polar orbits at an altitude between 450 and 250 km and with a separation of 100 to 200 km exchange microwave signals ("K-band signals" at K and Ka band) and simultaneously receive GPS L-band signals. The primary observables for gravity estimation are μm-level intersatellite biased range and its derivatives obtained from the phase of the K-band signals. Biased range is obtained by means of an improved version of the well-known dual-1-way technique (e.g., see [1]). GPS signals received by zenith-looking antennas are analyzed to extract pseudorange and phase, primarily for use in accurate orbit determination and clock synchronization. In a complete fit, which is not treated here, the K-band intersatellite observables and GPS observables are simultaneously fit over many days (e.g., 13 to 26 days) to extract gravity coefficients (e.g., up to degree 150, order 150) and the orbits of the two satellites.

This report introduces a dual-1-way approach for extracting "high-rate" K-band intersatellite biased range; develops a nonrelativistic model for measured biased range and selected errors; proposes improved filtering for extracting biased range, range rate, and range acceleration at a reduced rate from high-rate biased range; develops models for the random noise spectra for system noise and ultrastable-oscillator (USO) noise in the observables of range, range rate, and acceleration; and develops for purposes of illustration, intuition, and order-of-magnitude error analysis a hypothetical "fitting" approach based on estimating gravity coefficients from a simplified range-acceleration observable. Based on this crude acceleration approach, gravity-coefficient errors are generated by propagating observable errors to gravity-coefficient errors by means of both closed-form approximation and multiparameter sequential estimation. The closed-form approximations, which consist of a few simple equations, provide very fast, useful estimates of gravity-coefficient errors and insight into gravity estimation. The closed-form approximations require only seconds to run on a PC compared to a day for a full simulation on more powerful computers. Error analysis focuses here on three important random noise sources: USOs, accelerometers, and K-band system (background and thermal) noise and leaves a large assortment of other important errors for future reports.

Even though hypothetical and crude, the acceleration approach developed here for error estimation has introduced a number of useful insights and results related to the propagation of observable errors to gravity-coefficient errors, including the following:

a) improved filtering was developed for extracting the intersatellite observables of biased range, range rate, and range acceleration, including a new requirement limiting gravity-signal amplitude distortion;

b) the approach demonstrated the strengths of frequency-domain analysis in evaluating the fit filter for a given gravity coefficient and understanding error propagation to that coefficient;

c) the fit filter for estimating a given gravity coefficient can be approximately modeled in the frequency domain as a combination of lobes located at particular (gravity-coefficient-specific) tone frequencies, where each lobe has a very narrow bandwidth of 1 μHz (for a 13-day fit interval);

d) with a suitable fit model for instrumental effects, the random observable noise propagated to a given gravity coefficient is predominantly the noise found within the narrow bandpasses surrounding the aforementioned fit-filter lobes, mapped for each lobe according to lobe peak magnitude;

e) an observable error that is a tone, very nearly synchronous in frequency and phase with a given tone in the partial for a given gravity coefficient, is not attenuated in proportion to the very narrow lobe bandwidth, unlike random noise, but is passed according to the peak magnitude of the corresponding fit-filter lobe and therefore must be bounded by a separate error-tone specification;

f) a frequency-domain fit-filter method was developed for approximately assessing the effectiveness and consequences of using a piecewise-polynomial function of time to remove $f^{-\alpha}$ instrumental noise, a method that results in a number of useful insights and results, including the selection of polynomial order and update-interval length, the "mapping" of tones in a partial for a given gravity coefficient to lobes in the corresponding frequency-domain fit filter for that coefficient, gravity-coefficient error amplification due to excessive fit-filter mainlobes and sidelobes, and "separation" of orbit parameters from gravity coefficients in multiparameter fits;

g) an analysis was carried out to approximately assess the importance of correlations between estimated parameters, suggesting, in the case of the hypothetical acceleration fits, that correlations cause formal-error amplifications of approximately three or less, that correlations do not greatly increase per-degree geoid errors, and that cos(lat) weighting can offer substantial advantages;

h) based on the preceding results and on spectral models for observable noise, simple closed-form approximations were derived for crudely estimating gravity-coefficient errors for the hypothetical acceleration fits;

i) simulations based on the hypothetical acceleration approach, spectral models for observable noise and sequential multiparameter estimation, were used to propagate numerically simulated random noise from observables to gravity coefficients (up to degree 50, order 50); and

j) said simulations and closed-form analysis suggested that random per-degree geoid errors on the order of 10 μm (degree ≤ 50, 300-km altitude, 200-km range, 13-day fit) are feasible with respect to anticipated random errors due to system noise, USOs, and accelerometers.

The version of dual-1-way biased ranging developed by this report improves upon earlier versions in a number of ways: highly accurate satellite-timetag corrections derived from concurrent GPS data, better baseband phase extraction using highly digital processing, highly accurate USO-rate calibration derived from concurrent GPS data, an improved method for extracting high-rate biased range from baseband phase, improved filtering for extracting reduced-rate observables from high-rate biased range, and parallel extraction of three types of reduced-rate observables, namely biased range, range rate, and range acceleration. Satellite-clock synchronization from GPS data is expected to provide "absolute" synchronization at the 1- to 100-ns level, depending on the care

13

taken in calibrating instrumental delays, and time-varying synchronization errors at the 30-ps level [Willy Bertiger and Larry Young, private communications, 1998]. By sampling the baseband signals at a high rate (≈ 19 MS/s) and extracting phase with a digital phase-locked loop, one can reduce phase-extraction errors to very low levels. Periodic recalibration of USO rates on the basis of concurrent GPS data is expected to reduce the corresponding errors in gravity coefficients to less than a μm, per degree. The improved filtering proposed in this report for extraction of the reduced-rate observables from high-rate biased range leads to much smaller observable random errors and/or much less gravity-signal amplitude distortion (gain ripple as a function of frequency) than obtained with the prior dual-1-way methods considered in Appendix A. For example, based on the instrumental "performance levels" (i.e., level of input phase noise) assumed in this report for system noise and USOs, the proposed filtering improves range-rate random errors relative to a standard 10-s linear-least-squares fit by a factor of 10 to 35 in the important gravity-signal frequency range of 1 cyc/rev to 30 cyc/rev (see Appendix B), given an observable output rate of 0.1 sample/s (S/s). Gravity-signal amplitude distortion is reduced from the unacceptable levels found in the indicated prior methods to a negligible level by changing to the improved filtering. For example, at a signal-band frequency of 0.02 Hz (≈ 100 cyc/rev), fractional amplitude distortion is reduced from -0.04 for a 10-s linear fit to 10^{-5} for the improved filtering. Distortion of 10^{-5} at 0.02 Hz corresponds to a per-degree geoid error on the order of 1 μm or less. As a second example of earlier dual-1-way processing, the approach ("triangular weighting" option) proposed in Reference [1] for extraction of range rate can be implemented so that it leads to reasonable signal-band noise but it would introduce unacceptable gravity-signal amplitude distortion.

The improved filtering reduces the random errors for the range-acceleration observable by an even larger factor relative to a 10-s quadratic-fit with a 0.1-S/s output rate, by a factor of 10 to 1000 for the assumed performance levels for USO and system noise (see Appendix B). As a result, the acceleration observable can have a precision (i.e., random error) and amplitude distortion that are competitive with range and range rate. Because the acceleration observable can be analyzed, under certain crude approximations, to estimate gravity-coefficient errors without integrating the equations of motion, that observable has been selected for analysis in this report. It is emphasized that the hypothetical acceleration approach developed here for order-of-magnitude error analysis is not intended as a method for fitting actual measured accelerations to obtain gravity coefficients. The utility of the acceleration observable, if any, in the practical extraction of gravity information has yet to be demonstrated.

Earlier geoid-error estimates (early 1996 and before) typically analyzed the range-rate observable and effectively assigned a "white-noise" spectral error on the order of 3 to 8 $\mu m\ s^{-1}\ Hz^{-1/2}$ (i.e., 1 μm/s 1-σ noise per range-rate point given 10-s to 60-s sample spacing) [Srinivas Bettadpur, private communication, 1998]. In view of the results obtained by MacArthur and Posner [1] in 1985 and in light of the results of this report, these earlier geoid-error simulations adopted very conservative values for range-rate noise. Said adopted range-rate noise was too conservative by a factor of 40 relative to the 0.024-μm/s value "ideally" estimated by MacArthur and Posner for USO and system noise on the basis of "triangular weighting" and 10-s averages, and by a factor of 100 to 800 relative to the spectral error of 0.01 to 0.03 $\mu m\ s^{-1}\ Hz^{-1/2}$ estimated by this report in the important 1-to-30 cyc/rev gravity-signal band on the basis of the improved observable extraction and revised instrumental performance (see USO and system noise in Appendix B). One can show that the current specification for accelerometer noise also results in effective integrated range-rate noise far smaller than 1 μm/s over the important gravity-signal frequencies. Furthermore, these earlier geoid-error estimates did not model the strikingly nonuniform range-rate spectra from the various noise sources. These considerations indicate that these earlier simulations were too conservative with regard to their adopted levels for range-rate noise and emphasize the importance of refining noise simulations to model spectral shape.

Additional analysis and simulations have been carried out to estimate the geoid errors caused by observable noise generated by the improved observable extraction of this report, including simulation of noise spectral shape. Based on the new observable-noise models, the random gravity-coefficient errors per degree estimated in this work on the basis of the acceleration observable, improved by one or two orders of magnitude relative to the indicated prior simulations (e.g., see [2]), suggesting that per-degree geoid errors on the order of 10 μm (given aforementioned orbits, etc.) are feasible with regard to anticipated noise from the three modeled random error sources (system noise, USOs, accelerometers). The large improvement in random geoid errors obtained by this acceleration-observable analysis made it possible, with regard to the indicated random errors, to propose measurement of the relatively small time-varying components of the gravity field, which greatly increased the potential value of the GRACE mission. The greatly improved gravity-coefficient errors obtained here with the crude acceleration approach have been subsequently corroborated (order-of-magnitude) in the case of USO phase noise by full simulations based on the range-rate observable, after those simulations were revised to incorporate the improved observable-noise model of this report as well as other improvements [Srinivas Bettadpur, private communication, 1998]. These encouraging results pertain to the indicated random errors; the GRACE team has been analyzing a large assortment of other important error sources to estimate the corresponding geoid errors.

Section 2 outlines satellite instrumentation and signal processing and introduces the approach for extracting the observables of range, range rate and range acceleration from baseband phase. Section 3 presents the design and analysis for processing in more detail and introduces a theory for the intersatellite observables by tracing the signal from initial K-band transmission through the following steps: down-conversion of the received K-band signals at each satellite to baseband; sampling of the baseband signals at 19 MS/s, extraction of baseband phase with a digital phase locked loop; correction of clock-synchronization errors in timetags (based on GPS measurements); periodic recalibration of USO rates (based on GPS measurements); dual-1-way combination of measured phase to obtain "high-rate" (e.g., 10 S/s) intersatellite biased range; dual-band combination to correct for the ionosphere effect; and digital filtering of "high-rate" ionosphere-corrected biased range to obtain biased range, range rate, and range acceleration at a reduced rate (e.g., 0.1 S/s). A number of errors are formally modeled, including timetag errors, USO systematic and random phase variations, instrumental delay and phase errors, ionosphere effects, and system noise. This processing produces highly precise observables for intersatellite range (1 to 20 $\mu m\ Hz^{-1/2}$), range rate (0.01 to 0.1 $\mu m\ s^{-1}\ Hz^{-1/2}$), and range acceleration (10^{-4} to 0.01 $\mu m\ s^{-2}\ Hz^{-1/2}$), where the indicated spectral errors are due to system noise and USO noise and are a function of frequency in the gravity-signal band (e.g., 0.1 to 18 mHz up to degree 100). A theoretical expression is derived for the frequency-domain filtering that is applied to USO phase instabilities by the dual-1-way ranging method. An approximate theoretical model for the geometric component of the range, range-rate, and range-acceleration observables is presented and the correction of the acceleration observable for nongravitational accelerations is discussed.

For purposes of crude error analysis, Section 4 develops a simplified hypothetical fitting approach for extracting gravity coefficients from the acceleration observable, given known orbits. Based on the approach in Section 4, Section 5 develops a simple closed-form approximation for error propagation from acceleration observable to gravity coefficients. As a more accurate alternative to the closed-form approximation, Section 6 uses multiparameter sequential estimation based on the hypothetical approach to propagate numerically simulated system noise and accelerometer noise from the acceleration observable to gravity coefficients, simultaneously estimating gravity coefficients and instrumental parameters.

A number of appendices expand on specialized topics of importance. Appendix A demonstrates the suboptimality of the noise propagation and/or gravity-signal amplitude distortion resulting from a standard quadratic fit and other methods in extracting range, range rate, and

15

acceleration and then develops the aforementioned improved observable-extraction filtering and provides examples of filters. Appendix B presents examples of random noise spectra for system noise, USOs, and accelerometers and analyzes the frequency-domain filtering applied to system noise and USO noise by the improved signal processing in extracting range, range rate, and acceleration. Appendix C introduces a heuristic frequency-domain fit-filter method for assessing the effectiveness and consequences of using piecewise polynomial models to reduce $f^{-\alpha}$ instrumental noise. This analysis supports the zero-correlation and noise-mapping approximations of Section 5 and the modeling approach of Section 6, and introduces a number of useful fit-filter insights. Appendix D assesses the importance of correlations between gravity coefficients in the hypothetical acceleration fits. This analysis supports the zero-correlation approximation of Section 5 and suggests that interparameter correlations do not cause large increases in the per-degree geoid errors, with respect to parameters estimated in this report. The analysis does suggest, however, that interparameter correlations can substantially increase gravity-coefficient errors for some coefficients (by up to approximately a factor of three for white input noise for the coefficients tested). To illustrate the behavior of Legendre polynomials (LPs), Appendix E presents plots of various LPs and their derivatives.

SECTION 2

OUTLINE OF INSTRUMENTATION AND OBSERVABLE EXTRACTION

2.1 Satellite Instrumentation and Signal Processing

Fig. 2-1 presents a schematic block diagram of the intersatellite link and signal processing in the two GRACE satellites, including GPS processing. As suggested in this figure, the two satellites each will carry the following components: a K/Ka-band microwave antenna for transmitting and receiving the intersatellite signals; an upward-looking L1/L2-band microwave antenna for receiving GPS signals; a JPL-developed BlackJack GPS receiver for processing both the GPS signals and the down-converted K/Ka signals; a down-converter for the K/Ka intersatellite signals; a transmitter for generating the K/Ka signals for transmission; and an ultrastable oscillator (USO) to drive both the GPS and K/Ka systems. In addition, each satellite will carry a high-accuracy three-axis accelerometer for measuring nongravitational forces on the two GRACE satellites.

As schematically shown in Fig. 2-1, a single horn serves as the Ka/K (32/24 GHz) antenna for both transmitting and receiving the intersatellite dual-band μ-wave signals. These horns, which are mirror images of one another on the two satellites, are based on feed horns used in JPL's Deep Space Network (DSN). Each satellite transmits two sinusoidal signals (K and Ka). At one of the two satellites, the transmitter applies to each transmitted signal a frequency offset (nominally set to 0.5 MHz for illustration in this report) relative to the like-band signal transmitted by the other satellite, in order to separate the signals. At each satellite, each received K-band signal is down-converted to 0.5 MHz by using as a local-oscillator (LO) signal the like-band sinusoidal signal (i.e., at K or Ka) to be transmitted by that satellite. The two 0.5-MHz down-converted RF signals are each sampled at approximately 19 MHz and passed to the digital signal processing (DSP) part of the BlackJack receiver. Dedicated channels of BlackJack are used to digitally counter-rotate the phase of each down-converted signal, track phase with a digital phase-locked loop, and extract phase. For each band (K and Ka), BlackJack outputs phase in cycles at a selected rate (nominally 10 samples per second for this report), and these values are transmitted to the ground for further processing.

At each satellite, a zenith-looking L1/L2 antenna collects the signals transmitted by concurrently visible GPS satellites. BlackJack hardware and software simultaneously process the signals for as many as eleven GPS satellites to produce highly-accurate pseudorange and carrier phase measurements for each GPS signal. These measurements are transmitted to the ground where they are processed (along with GRACE intersatellite observables and GPS observables from a global network of receivers) to assist in obtaining accurate orbits and clock synchronization for the two GRACE satellites.

At each satellite, the USO is the primary frequency standard for radiometric components, driving the transmitted K/Ka-band signals, K/Ka-band LO signals, GPS LO signals, and samplers for both GPS and GRACE signals. This design provides ultrastable phase to radiometric parts of the system and ultimately leads to the added instability reduction (by a factor on the order of 10^4 to 10^5) provided by dual-1-way ranging when the intersatellite phase values measured by the two satellites are combined on the ground.

17

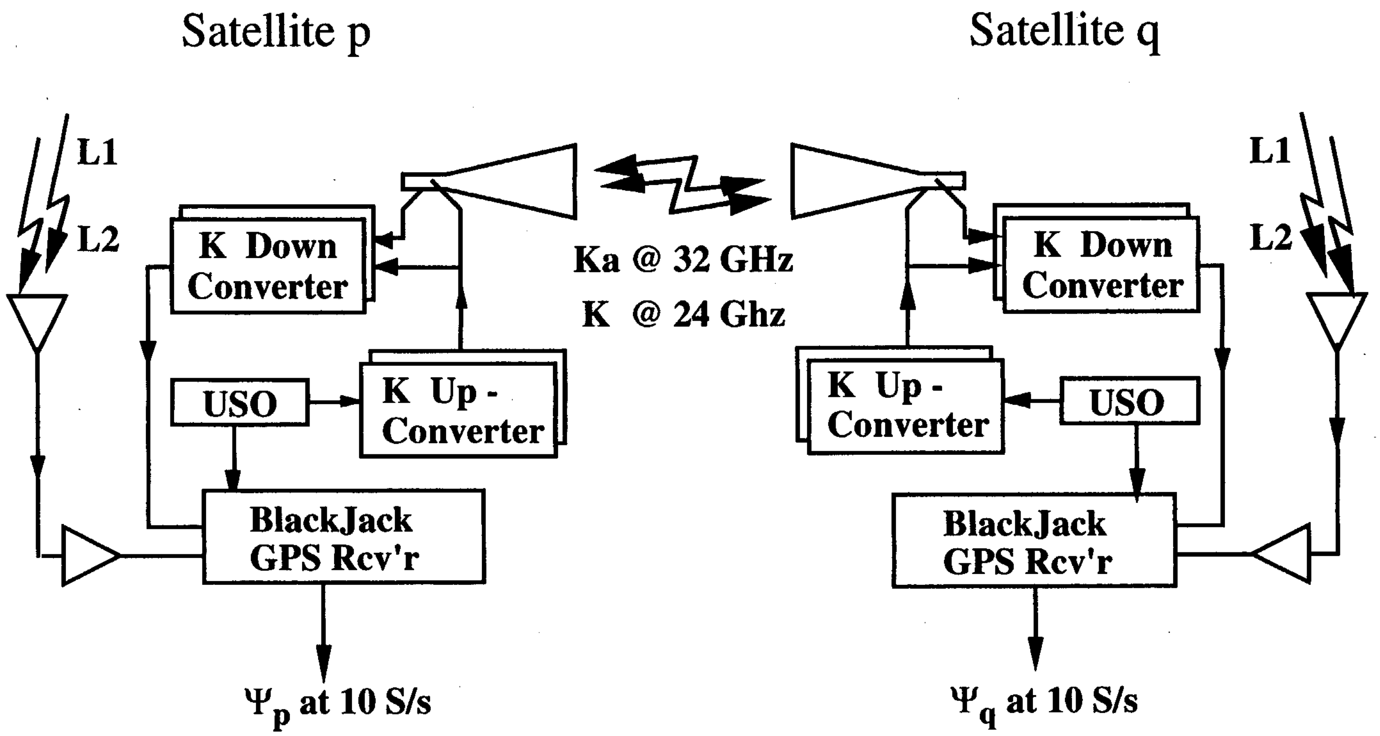

Figure 2-1 Schematic Block Diagram for Integrated Microwave Link and GPS Receivers

2.2 Extraction of the Observables of Range, Range Rate and Acceleration

The phase values, separately extracted for the K and Ka bands in units of cycles by BlackJack, are transmitted to the ground by each satellite at the nominal rate of 10 S/s for further processing. (In this report, S/s denotes samples per second.) Fig. 2-2 summarizes an example implementation for extracting the ionosphere-corrected observables of (biased) range, range rate and range acceleration. The processing outlined here does not necessarily represent the optimal implementation for actual processing but is presented to illustrate important variables and concepts through the use of example processing.

Based on timetag offsets derived from GPS data for each satellite, the sample time of the phase from each satellite is separately adjusted to correct, to some level of accuracy, for timetag error (see Section 3). The phase values from the two satellites at the same nominal corrected time point are then added, separately for K and Ka bands, and the result is multiplied by the speed of light and divided by the sum of the like-band carrier frequencies (e.g., by 24 GHz + (24 GHz + 0.5 MHz) for K band). This operation generates separately for K and Ka band "dual-1-way range" values ("biased range") at the nominal rate of 10 S/s, with greatly reduced errors due to long-term drifts of the USOs. (A nominal rate of 10 S/s is adopted here for illustration.)

The next processing step applies the standard dual-band linear combination of K and Ka range values (with the same timetag) to remove ionospheric effects. The output of this processing step is ionosphere-corrected (biased) range and ionosphere delay, each at a nominal rate of 10 S/s.

The next processing step extracts range, range-rate and acceleration observables from the 10-S/s range values and decimates the output data rate to a selected value, which is set to 0.1 S/s in this example. (Actual processing is likely to involve a choice of output rate, tailored to the user's needs, as discussed in Subsection 3.1.5) Appendix A explains in detail the selection of the method used to carry out this observable-extraction, rate-reduction step. The selected method must filter noise in an optimum manner and not distort the amplitudes of gravity harmonics across the signal band. As explained in Appendix A and illustrated in Fig. 2-2, the 10 S/s biased range values for each band are passed through three parallel non-recursive time-domain digital filters: one for generating biased range, one for range rate, and one for range acceleration. They each simultaneously apply a low-pass filter (e.g., 0.05-Hz), reduce the data rate (e.g., by a factor of 100 from 10 S/s to 0.1 S/s), and, for range rate and acceleration, differentiate the range values to the appropriate order . Note that a 0.05-Hz bandwidth for the low-pass filter is consistent with an output rate of 0.1 S/s so that high-frequency, out-of-band noise cannot alias into the sampling band (0 to 0.05 Hz). The "gain ripple" of the filters in this example is constructed so that amplitude is adequately preserved for tones up to the maximum frequency in the selected gravity-signal band (e.g., for an output rate of 0.1 S/s, approximately 0.1 to 25 mHz for harmonics up to degree 135). The output of this step contains three data types: biased range, range rate, and range acceleration, each at the selected output rate (e.g., 0.1 S/s). Any one or any combination of these final observables can potentially be used in gravity analysis.

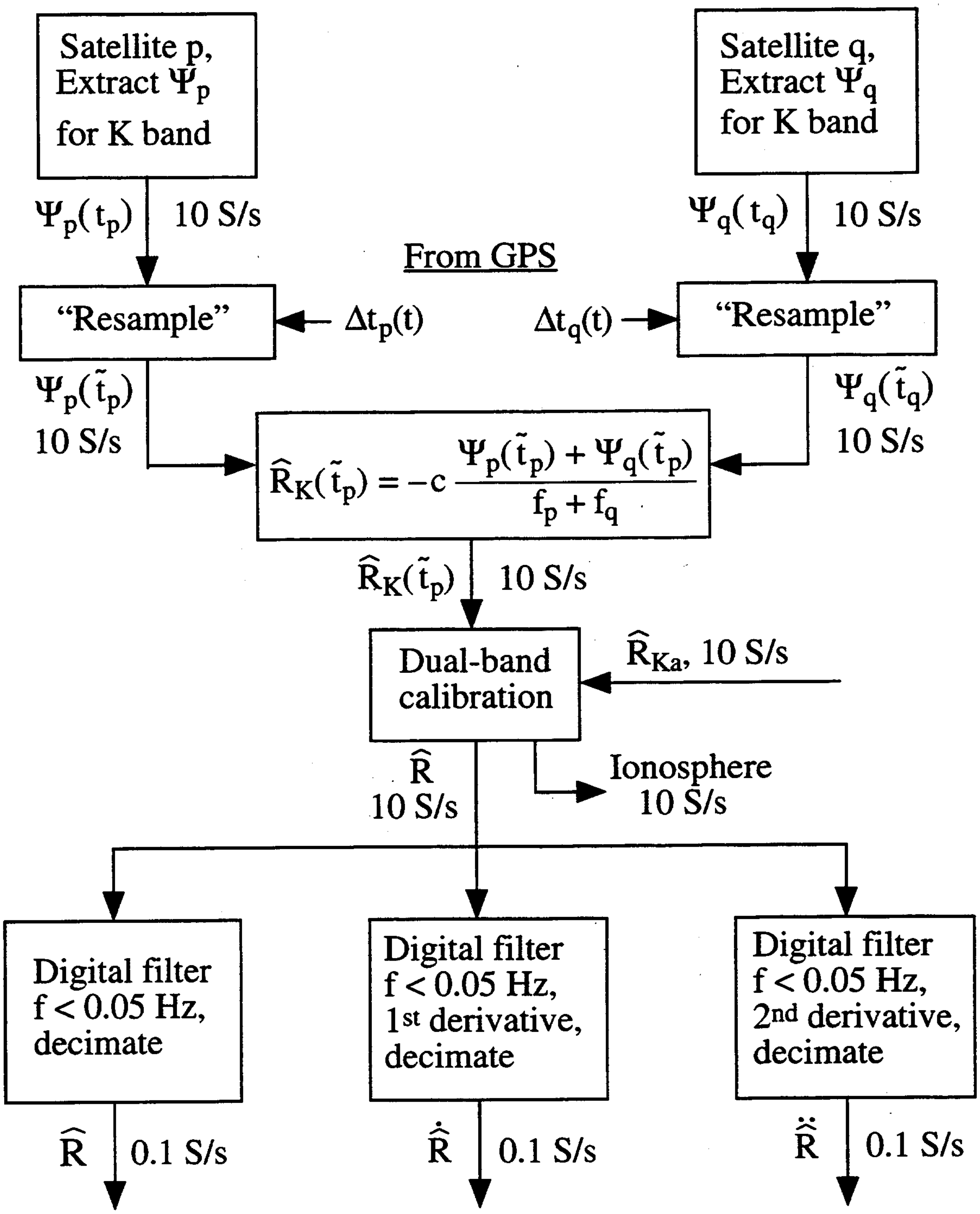

Figure 2-2 Schematic Example of Processing Steps to Extract Intersatellite Observables

SECTION 3

THEORY FOR INTERSATELLITE OBSERVABLES

This section analyzes the intersatellite observables and models observable extraction, the geometric component of the observables, and the nongravitational correction to the acceleration observable. The analysis is approximate in that a number of features and error sources are not treated, including but not limited to: relativistic effects, multipath, mapping from K-band-antenna phase center to satellite center of gravity (CG), offset of satellite CG from accelerometer-proof-mass CG, angular misalignment of satellite pointing and accelerometer, systematic in-signal-band USO instabilities, systematic in-signal-band accelerometer errors, accelerometer-scale-factor calibration, accelerometer gain nonlinearities, neutral atmosphere, and ionosphere-calibration errors. This report focuses on the random errors of system noise, USO noise, and accelerometer noise. The GRACE team has been analyzing the other important error sources to estimate the corresponding errors in geoid measurements.

3.1 A Model for the Extraction of the Intersatellite Observables

The primary features of the measured intersatellite observables can be demonstrated with the following approximate nonrelativistic model that traces phase and its errors and instabilities through the stages that lead to the observables. Figs. 2-1 and 2-2 present the system at a high level, while Fig. 3-1 presents a more detailed schematic illustration of satellite processing, as designed by the GRACE team.

3.1.1 A Model for Sampled Baseband Phase

This subsection develops a nonrelativistic theoretical model for the baseband phases extracted at each satellite for K band. A parallel analysis, which is not presented here, can be carried out for Ka band. As observed at each respective transmitting satellite, let $\phi_p(t)$ and $\phi_q(t)$ be the phases of the sinusoidal signals (e.g., $A_T \cos[\phi_p(t)]$ for satellite p) transmitted at (common) true time t by satellites p and q, respectively, for K band at RF frequencies given by $f_p \approx 24$ GHz for satellite p and $f_q \approx 24$ GHz + 0.5 MHz for satellite q. These transmitted phases, in cycles, can be modeled at RF as

$$\phi_p(t) = \alpha_p(t) + \eta_p(t) \tag{3.1}$$

and

$$\phi_q(t) = \alpha_q(t) + \eta_q(t) \tag{3.2}$$

where α_p and α_q represent long-term trends (e.g., > 20,000 s), including the linear term (e.g., 24 GHz), while η_p and η_q represent short-term variations (random or tone-like), which are of particular concern when they are in the signal band (e.g., 0.1 to 36 mHz). To first approximation, α_p and α_q each consist of a linear term and a small quadratic (aging) term (e.g., $\approx f_o t + 0.5 \dot{f} t^2$ where $f_o \approx 24$ GHz and $|\dot{f}/f_o| \leq 5\text{x}10^{-10}$/day). In this report, η_p and η_q are restricted to random phase deviations (e.g., USO phase noise) while short-term systematic errors (e.g., due to temperature, radiation, and the Earth's magnetic field) are neglected.

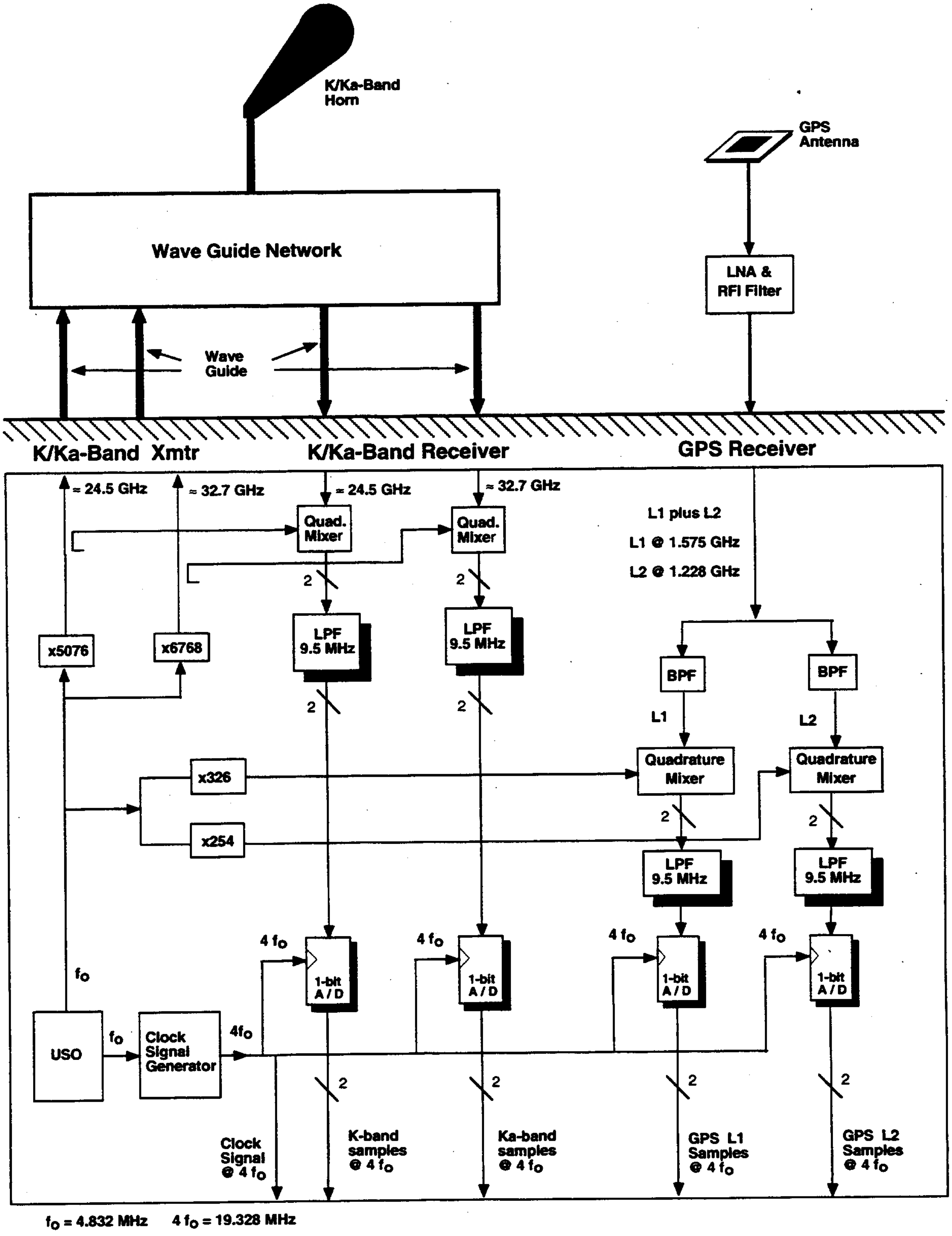

Figure 3-1 Schematic Illustration of Signal Processing on a GRACE Satellite

If $\tau_p(t)$ is the effective delay from satellite q to satellite p as "observed" with respect to satellite-p frame, including all effects, the phase of the received sinusoidal signal observed at satellite p at true time t is approximately $\phi_q(t - \tau_p) + I_q$, where I_q is the phase shift due to the ionosphere at frequency f_q. (In terms of transmit/receive times, t is the receive time and $t - \tau_p$ is the transmit time.) If $\tau_q(t)$ is the effective delay from satellite p to satellite q as "observed" with respect to satellite-q frame, the phase of the received sinusoidal signal observed at satellite q at true time t is approximately $\phi_p(t - \tau_q) + I_p$, where I_p is the phase shift due to the ionosphere at frequency f_p. Let the effective intersatellite delay from satellite q to satellite p be represented as

$$\tau_p(t) = \tau(t) + \delta\tau_p \tag{3.3}$$

and from satellite p to satellite q as

$$\tau_q(t) = \tau(t) + \delta\tau_q \tag{3.4}$$

where $\tau(t)$ is the geometric delay ("light-time delay") at true time t, including satellite motion during signal transit, and where $\delta\tau_p$, $\delta\tau_q$ include any effects (e.g., instrumental or neutral atmosphere) that can cause effective delay to differ from the light-time delay. As a result of symmetry, note that, for the same receive (true) time t at both satellites, the light-time delay from satellite q to satellite p is the same as the light-time delay from satellite p to satellite q.

In practice, the light-time delay between satellites refers to the delay from phase center to phase center. Corrections and/or errors related to mapping from phase center to satellite center of mass are not analyzed in this report. Because of tight attitude control that places both phase centers very close to the line between satellite centers of mass, the measured time variations between phase centers is very close to the time variations between centers of mass.

Based on the preceding discussion, the received signal at satellite p at true time t can be modeled as

$$V_R(t) = A_R \cos[\phi_q(t - \tau_p) + I_q] \tag{3.5}$$

where A_R is the received signal (voltage) amplitude. As indicated in Fig. 3-1, the transmitted phase at each satellite is used to down-convert the incoming signal so that, in effect, $\phi_p(t)$ is subtracted from the received phase at satellite p and $\phi_q(t)$ is subtracted from received phase at satellite q. This down-conversion is carried out in quadrature (i.e., to produce cosine and sine components) through use of a quadrature mixer. For example, given an "ideal" quadrature mixer at satellite p, one can model the mixer as a device driven by two inputs, $\cos(\phi_p)$ and the received RF signal V_R, which the mixer uses to generate two outputs, the products (quadrature "channels") $V_R \cos(\phi_p)$ and $V_R \sin(\phi_p)$. Since each of these products is a sinusoid times a sinusoid, each can be expressed as the sum of two sinusoids, one with phase equal to the sum (sum note) of the received phase and ϕ_p, and the other with phase equal to the difference (beat note). Each mixer output channel is passed through a filter that removes the sum note, leaving only the (baseband) difference note, which can be modeled as a cosine function for one channel and a sine function for the other.

Thus, at each satellite, the baseband signal consists of two quadrature sinusoidal voltage components. For satellite p, for example, the cosine baseband component can be modeled as

23

$$V_c(t) = A_B \cos[\phi_q(t - \tau_p) - \phi_p(t) + I_q + \gamma_p] + \eta_{\cos} \tag{3.6}$$

and the corresponding sine component as

$$V_s(t) = A_B \sin[\phi_q(t - \tau_p) - \phi_p(t) + I_q + \gamma_p] + \eta_{\sin} \tag{3.7}$$

where A_B is the baseband-signal amplitude, and $\eta_{\cos}$ and $\eta_{\sin}$ are random amplitude noise (see system noise in Appendix B) contributed by instrumentation and background noise. The phase term γ_p has been added to account for phase errors contributed by instrumentation and processing at satellite p, including potential phase ambiguities. Corresponding expressions represent the baseband signals for satellite q.

At each satellite, each quadrature baseband component modeled by Eqs. (3.6) and (3.7) is separately passed through a 9.5-MHz lowpass filter and then sampled at 19.328 MS/s (to be referred to as 19 MS/s), with timetags assigned according to the satellite clock. Phase is then extracted from the 19-MS/s quadrature samples at a selected rate (nominally 10 S/s) using a digital phase-locked loop (DPLL, e.g., see [3] and [4]), with an overall integer-cycle phase ambiguity. The effective filter applied to the phase values by this processing is a sinx/x filter (with a first null at 10 Hz for the nominal 10 S/s sample rate). Amplitude distortion introduced by this filter is minimal, reaching only 2.6×10^{-5} at the worst frequency point (i.e., at the upper edge of the signal band where amplitude is reduced by $\sin\phi/\phi$ for $\phi = \pi * 0.04$ Hz/10 Hz, assuming a maximum signal frequency of 0.04 Hz). This amplitude distortion is to be compared with the nominal specification of 3×10^{-3} at 0.04 Hz (see Appendix A).

For each satellite, the timetags that are assigned by the sampler according to the satellite clock are offset from true time:

$$t_p = t + s_p(t) \tag{3.8}$$

$$t_q = t + s_q(t) \tag{3.9}$$

where t_p and t_q are sampler time for satellites p and q, respectively. The terms s_p and s_q are the respective synchronization offsets of satellite time from true time, each including (but not limited to) a bias, a USO rate offset and USO noise. For satellite p, baseband phase, which is modeled as $\phi_q(t - \tau_p) - \phi_p(t) + I_q + \gamma_p$ in Eqs. (3.6) and (3.7), is extracted by the DPLL at a 10-S/s rate from the 19-MS/s baseband samples. The extracted phase values, which are each labeled with a timetag provided by the satellite clock, will be denoted as $\Psi_p(t_p)$ at satellite time t_p for satellite p and modeled as

$$\Psi_p(t_p) = \phi_q(t_p - s_p - \tau_p) - \phi_p(t_p - s_p) + I_q + \gamma_p + \zeta_p \tag{3.10}$$

where theoretical true time t found in the phase arguments of Eqs. (3.6) and (3.7) has been replaced by $t_p - s_p$ in order to express theoretical phase in terms of the known assigned timetag t_p. Each timetag is at the center of its 0.1-s interval. The corresponding expression for satellite-q phase is given by

$$\Psi_q(t_q) = \phi_p(t_q - s_q - \tau_q) - \phi_q(t_q - s_q) + I_p + \gamma_q + \zeta_q \tag{3.11}$$

24

Even though not explicitly shown, the intersatellite delays, τ_p and τ_q, become functions of time with the form $\tau_p(t) = \tau_p(t_p - s_p)$ and $\tau_q(t) = \tau_q(t_q - s_q)$, as indicated by Eqs. (3.3), (3.4), (3.8), and (3.9). Note that new phase errors, ζ_p and ζ_q, have been included to account for the system noise shown in Eqs. (3.6) and (3.7).

The 10-S/s phase values produced by the satellite DPLLs and modeled by Eqs. (3.10) and (3.11) are transmitted to ground for further processing.

3.1.2 Correction of Satellite-Clock-Synchronization Errors

This subsection expands on the resampling step in Fig. 2-2. The timetags, t_p and t_q, labeling the Ψ_p and Ψ_q samples at each satellite on the basis of nominal satellite-clock reading can deviate greatly from true time due to initial rate offsets of the USOs (with a specification of $|\Delta f/f| \leq 5\text{x}10^{-8}$) and USO aging (with a specification of $|\dot{f}/f|$ less than $5\text{x}10^{-10}$/day). For example, a rate offset of $\Delta f/f = 5\text{x}10^{-8}$ can cause a timetag shift of 4 ms over 1 day. One approach for handling such large timetag errors in ground processing is to "adjust" the sample times for the sampled values of Ψ_p and Ψ_q (i.e., to "resample") so that the resulting new timetags are closer to true time and one another, and are uniformly spaced (still at the nominal rate of 10 S/s) on the new time scale. The synchronization corrections, Δt_p and Δt_q, used to drive such resampling can be derived from concurrently observed GPS signals.

At each satellite, all samplers and down-converters, for both K-band and L-band, are driven by the same USO (see Fig. 3-1). With such commonality, concurrent measurements of phase and pseudorange to each visible GPS satellite can be analyzed to accurately estimate the time-dependent offset of each satellite clock relative to an accurate "reference time" established through the GPS network. For the purposes of this report, this reference time is adopted as "true" time t. The errors in such GPS-derived offsets, when extracted for a <u>ground-based</u> receiver, have been demonstrated to be on the order of 1 to 100 ns in absolute ("bias") terms relative to "reference time" (depending on the care taken in calibrating instrumental delays) and 30 ps in relative (i.e., time variation) terms [Willy Bertiger and Larry Young, private communications, 1998]. These numbers may be better for GRACE satellites since the instrumentation in the satellites is highly stabilized, and clock offsets appear mostly as a difference between satellites (see Eqs. (3.19) and (3.20) below), which should provide some common-mode reduction of errors.

Suppose the time-dependent offset for a satellite clock relative to reference time has been estimated (say, every 10 s) on the basis of GPS data. For illustration in this analysis, two approaches are considered for using these offset values: a) applying "smoothed" (fit) values intended primarily to track longer-term (e.g., > 20,000 s) systematic USO errors (nominally to within 1 to 100 ns for the bias and to better than, say, 1 ns for long-term variations) or b) applying the "high-rate" values (e.g., 1/10-s) to correct for both short- and long-term variations to the accuracy allowed by GPS clock synchronization (nominally to within 1 to 100 ns for the bias error and 30 ps for time-varying error). The case a) approach would be used if the signal-band clock-synchronization variations caused by USO stability were smaller than the signal-band GPS-synchronization errors. The case b) approach would be used if GPS synchronization errors were smaller. The best approach will depend on a signal-band spectral comparison of the GPS clock-synch errors with the clock variations induced by USO instability. This might lead to an intermediate "smoothing" time for GPS synchronization measurements. Current estimates suggest GPS clock-synchronization errors might be smaller than USO-induced clock variations, at least at lower signal-band frequencies (e.g., 1 mHz and lower).

One approach for timetag "adjustment" is the following "resampling" method. Let $\Delta t_p(t)$ be the clock offset (smoothed or unsmoothed) estimated on the basis of GPS data for satellite p. Define a new, more accurate time scale, $\tilde{t}_p$, in terms of the old time scale, t_p, by the relation, $t_p = \tilde{t}_p + \Delta t_p(t)$. "Resampling" in terms of the new more accurate time scale can be accomplished by first defining uniformly spaced sample points in new time (nominally still at 10 S/s). For each uniformly-spaced value of $\tilde{t}_p$, one can use the relation, $t_p = \tilde{t}_p + \Delta t_p(t)$, to determine the corresponding time point on the t_p time scale. A phase value for this point is then computed by a suitable interpolation method on the basis of the original phase values that are uniformly spaced as a function of t_p. A corresponding operation is separately carried out for t_q for satellite q.

When properly carried out, this "resampling" operation amounts to a simple time shift that results in the phase samples that would have been generated by a sampler with the improved time synchronization. Based on Eq. (3.10), the result can be modeled by simply substituting $\tilde{t}_p$ for t_p in the argument of Ψ_p on the left-hand side to denote the new timetag label and replacing t_p with $\tilde{t}_p + \Delta t_p(t)$ on the right-hand side, with corresponding operations for Eq. (3.11). After such resampling, the measured baseband phases from the two satellites can be modeled as

$$\Psi_p(\tilde{t}_p) = \phi_q(\tilde{t}_p - \tilde{s}_p - \tau_p) - \phi_p(\tilde{t}_p - \tilde{s}_p) + I_q + \gamma_p + \zeta_p \tag{3.12}$$

and

$$\Psi_q(\tilde{t}_q) = \phi_p(\tilde{t}_q - \tilde{s}_q - \tau_q) - \phi_q(\tilde{t}_q - \tilde{s}_q) + I_p + \gamma_q + \zeta_q \tag{3.13}$$

in which $\tilde{s}_p \equiv s_p - \Delta t_p$ and $\tilde{s}_q \equiv s_q - \Delta t_q$ and in which the implicit timetags for τ_p, τ_q, I_p, I_q, γ_p, and γ_q are also shifted. For example, the delay τ_p is evaluated at $\tilde{t}_p - \tilde{s}_p$ and the delay τ_q is evaluated at $\tilde{t}_q - \tilde{s}_q$. The new timetag errors, $\tilde{s}_p$ and $\tilde{s}_q$, are much smaller than the original errors, as indicated above.

3.1.3 Extraction of Intersatellite Biased Range

This subsection develops a nonrelativistic theoretical model for the dual-1-way biased range extracted for K band in Fig. 2-2 from Ψ_p and Ψ_q. Each measurement of biased range $\hat{R}_K$ (to be called "range" for brevity) is computed by adding the two baseband phase values measured at the same corrected receiver time at the two satellites (i.e., $\tilde{t}_p = \tilde{t}_q$), changing sign, dividing by the sum of estimated USO RF rates, and multiplying by the speed of light:

$$\boxed{\hat{R}_K(\tilde{t}_p) \equiv c\hat{\tau} \equiv -c\,\frac{\Psi_p(\tilde{t}_p) + \Psi_q(\tilde{t}_p)}{f_p(\tilde{t}_p) + f_q(\tilde{t}_p)}} \tag{3.14}$$

Even though the estimates for USO rates (f_p and f_q) in the denominator are slowly changing ($|\dot{f}/f| < 5\text{x}10^{-10}$/day), they have been given time arguments to emphasize the necessity of regular re-estimation, as discussed below. The particular function of phases given in Eq. (3.14) has been selected since it leads to a result equal to c times the light-time delay between satellites ($c\tau$, to be referred to as "light-time range") plus error terms, with the advantage that the error terms due to

USO instability are reduced by self-cancellation, as shown below. Note that the forced assumption, $\tilde{t}_p = \tilde{t}_q$, results in differential timetag errors, as shown below in Eqs. (3.19) and (3.20). That is, if one enforces the relation $\tilde{t}_p = \tilde{t}_q$, then the true time difference is $\tilde{s}_p - \tilde{s}_q$. Even though it is not explicitly displayed, the following derivation implicitly accounts for this difference in true times.

When Eqs. (3.12) and (3.13) are substituted in Eq. (3.14), one obtains

$$\hat{R}_K(\tilde{t}_p) =$$

$$-c\,\frac{\phi_p(\tilde{t}_p - \tilde{s}_q - \tau_q) - \phi_p(\tilde{t}_p - \tilde{s}_p) + \phi_q(\tilde{t}_p - \tilde{s}_p - \tau_p) - \phi_q(\tilde{t}_p - \tilde{s}_q) + I_p + \gamma_q + \zeta_q + I_q + \gamma_p + \zeta_p}{f_p(\tilde{t}_p) + f_q(\tilde{t}_p)} \tag{3.15}$$

where the delay τ_p is evaluated at time $\tilde{t}_p - \tilde{s}_p$ and the delay τ_q is evaluated at time $\tilde{t}_p - \tilde{s}_q$. Thus, dual-1-way ranging causes the RF phase of each satellite to be subtracted from itself, with a small time offset on the order of the delay between satellites ($\tau \approx 1$ ms), provided the corrected-timetag errors are small.

When Eqs. (3.1) and (3.2) are substituted in Eq. (3.15), one obtains differences of the systematic phase terms and of the phase noise terms. The differences for the systematic terms, α_p and α_q, will be approximated by derivatives in order to reveal τ, while the differences for the noise terms, η_p and η_q, will be retained as differences for later analysis. The conversion to a derivative is accomplished by expansion about $\tilde{t}_p - \tilde{s}_p$. The expansion of the α_p terms about $\tilde{t}_p - \tilde{s}_p$ is given by

$$\alpha_p(\tilde{t}_p - \tilde{s}_q - \tau_q) - \alpha_p(\tilde{t}_p - \tilde{s}_p) \approx \dot{\alpha}_p(\tilde{t}_p - \tilde{s}_p)\,[\tilde{s}_p - \tilde{s}_q - \tau_q(\tilde{t}_p - \tilde{s}_p) + (\tilde{s}_q - \tilde{s}_p)\dot{\tau}] \tag{3.16}$$

if one assumes $\dot{\tau}_q \approx \dot{\tau}$. In this expression, $\tau_q(\tilde{t}_p - \tilde{s}_q)$ has been expanded about $\tilde{t}_p - \tilde{s}_p$, which results in the $\dot{\tau}$ term. The expansion of the α_q terms about $\tilde{t}_p - \tilde{s}_p$ is given by

$$\alpha_q(\tilde{t}_p - \tilde{s}_p - \tau_p) - \alpha_q(\tilde{t}_p - \tilde{s}_q) \approx \dot{\alpha}_q(\tilde{t}_p - \tilde{s}_p)\,[\tilde{s}_q - \tilde{s}_p - \tau_p(\tilde{t}_p - \tilde{s}_p)] \tag{3.17}$$

The neglected quadratic terms are approximately $0.5\ddot{\alpha}\tau^2$ which is on the order of 10^{-10} cyc or 10^{-6} μm, given accurate new timetags, a worst-case USO rate drift of $|\dot{f}/f| = 5\mathrm{x}10^{-10}$/day and $\tau = 1$ ms.

If one substitutes Eqs. (3.1), (3.2), (3.3), (3.4), (3.16), and (3.17) in Eq. (3.15), one obtains a theoretical expression for measured range for K band at corrected receiver time $\tilde{t}_p$:

$$\boxed{\hat{R}_K(\tilde{t}_p) \approx c\tau(\tilde{t}_p - \tilde{s}_p) + R_{tt} + R_{\dot{\tau}} + R_\phi + R_{\delta\tau} + R_I + R_\gamma + R_\eta} \tag{3.18}$$

As anticipated above, measured range has turned out to be equal to light-time range plus a number of error terms. The error terms include the following: an error in range due to the relative timetag error arising directly from the arguments of α_p and α_q, given by

$$R_{tt} \equiv c\,\frac{f_q - f_p}{f_p + f_q}\,(\tilde{s}_p - \tilde{s}_q)\,, \tag{3.19}$$

an error due to relative timetag error arising indirectly from the argument in τ_q (see Eq. (3.4) and (3.16)), by

$$R_{\dot{\tau}} = \frac{c\,\dot{\tau}\,f_p\,(\tilde{s}_p - \tilde{s}_q)}{f_p + f_q}\,, \tag{3.20}$$

an error due to RF phase noise by

$$R_\phi \equiv c\,\frac{\eta_p(\tilde{t}_p - \tilde{s}_p) - \eta_p(\tilde{t}_p - \tilde{s}_q - \tau_q) + \eta_q(\tilde{t}_p - \tilde{s}_q) - \eta_q(\tilde{t}_p - \tilde{s}_p - \tau_p)}{f_p + f_q}\,, \tag{3.21}$$

a delay-error term by

$$R_{\delta\tau} = c\,\frac{f_p\,\delta\tau_q + f_q\,\delta\tau_p}{f_p + f_q}\,, \tag{3.22}$$

an ionosphere error by

$$R_I \equiv -c\,\frac{I_p + I_q}{f_p + f_q}\,, \tag{3.23}$$

an instrumental phase error by

$$R_\gamma \equiv -c\,\frac{\gamma_p + \gamma_q}{f_p + f_q}\,, \tag{3.24}$$

and a system-noise error by

$$R_\eta \equiv -c\,\frac{\eta_p + \eta_q}{f_p + f_q}\,. \tag{3.25}$$

To obtain Eq. (3.18), it is assumed the time-specific RF USO rates, f_p and f_q, estimated on the basis of GPS data, are sufficiently close to actual rates. That is, $f_p + f_q \approx \dot{\alpha}_p + \dot{\alpha}_q$. For the largest gravity coefficient, J2, an error of a part in 10^9 in these USO rate estimates results in a geoid error on the order of 3 μmG. Other gravity coefficients lead to far looser requirements for USO rate errors. For a worst-case rate drift of $|\dot{f}/f| = 5\text{x}10^{-10}$/day, the fractional rate change would be $5\text{x}10^{-10}$ in a day. This result indicates re-estimation of USO rates for use in Eq. (3.14) could be carried out, say, once or twice per day. The error in the USO rate estimates derived from GPS data is expected to be orders-of-magnitude smaller than 10^{-10} [Willy Bertiger and Larry Young, private communications, 1998].

Thus, the measured range in Eq. (3.18) for the given band is equal to the light-time range, $c\tau$, between the satellites plus seven terms. The light-time delay to a given satellite can be viewed as $\tau = t_R - t_T$, where t_R is a selected receive (true) time at that satellite and t_T is the corresponding transmit time from the other satellite, including the transit-time effect. Relative to the receiving satellite, the speed of the transmitting satellite is on the order of 1 m/s so that the transit-time effect ($\approx v\tau$) is on the order of 1 m/s * 1 ms = 1 mm for measured range. The time argument for τ has been formulated as $\tilde{t}_p - \tilde{s}_p$, where $\tilde{s}_p$ is the absolute error in $\tilde{t}_p$ relative to reference (true) time. If $c\dot{\tau} \approx 1$ m/s and $\tilde{s}_p$ is less than 100 ns, the range error due to absolute clock-synchronization error is less than 0.1 μm.

The second term on the right of Eq. (3.18) accounts for the "direct" effect of timetag errors. A constant offset or slow systematic variation in timetag error (e.g., uncorrected rate drift due to USO aging) can be absorbed by the instrumental model in gravity estimation. The random component in this term that is contributed by USO phase noise (for case-a GPS synchronization corrections) is analyzed in Subsection 3.2. This term is a result of the RF frequency difference between satellites, $f_q - f_p$, which causes a large frequency offset (currently approximately 500 or 700 kHz) of equal magnitude in the sampled baseband signal at each satellite. This frequency offset leads to additional range error by allowing variations in sampler synchronization error to show up as phase variations. For a nearly zero baseband frequency (i.e., $f_q - f_p \approx 0$), sampler synchronization variations would have negligible effect, as Eq. (3.19) indicates. However, practical considerations prevent implementation of a value for $f_q - f_p$ close to zero. For example, the BlackJack receiver requires a nonzero baseband frequency in order for phase counter-rotation to work properly. Since the rates of both USOs can drift ("spec'ed" at $<5\times10^{-10}$/day or <30 kHz at Ka band over 5 years), a safe margin in USO rate offset (say, 100 to 200 kHz) should be implemented to prevent a zero crossing of $f_q - f_p$ due to USO drift. However, this zero-crossing requirement has been superseded by another requirement; an even higher specification of 500 kHz has been placed on $f_q - f_p$ due to concerns about "separation" in analog processing.

The third term, Eq. (3.20), accounts for the indirect effect of time tag errors that are in the arguments of τ. If one assumes that $c\dot{\tau} \approx 1$ m/s and that the GPS-based timetag correction outlined above can remove the timetag error between satellites to better than 100 ns (e.g., $|\tilde{s}_p - \tilde{s}_q| < 100$ ns, constant plus variable), then $|R_{\dot{\tau}}| < 0.1$ μm.

The fourth term, Eq. (3.21), accounts for instrumental RF phase noise, including, for example, USO noise and up-conversion noise. In a fit for gravity coefficients, an instrumental model (see Appendix C) is included to account for this error term, along with others. The random component of USO instability is analyzed in Subsection 3.2. Similar to 2-way ranging, in the dual-1-way technique, the phase of each USO is subtracted from itself with a small ($\tau \approx 1$ ms) time offset, as indicated by Eq. (3.21). As discussed in Subsection 3.2, such subtraction greatly reduces the effect of USO instability, by a factor on the order of 10^4 or larger, depending on parameter values. Note that this self cancellation applies not only to USO instability but also to instabilities introduced by common-mode up-conversion components between each USO (at ≈ 5 MHz) and RF. (Common mode here comprises all components between a USO and the point where the RF signal to be used for transmission separates from the signal to be used for down-converting the incoming signal.)

The fifth term, Eq. (3.22), accounts for any delays that can cause the effective delay between satellites to deviate from light-time delay. The sixth term, Eq. (3.23), accounts for ionosphere error, which is considered in the next subsection. The seventh term, Eq. (3.24), accounts for all other instrumental phase effects, including baseband phase shifts. Finally, the eighth term, Eq. (3.25), accounts for system noise errors, which are analyzed in Appendix B.

3.1.4 Dual-Band Ionosphere Calibration

This subsection summarizes the dual-band calibration operation applied in Fig. 2-2 to correct for ionospheric effects. To good approximation, the ionospheric phase shifts in Eq. (3.23) are inversely proportional to frequency:

$$I_p = C_I / f_p \tag{3.26}$$

and

$$I_q = C_I / f_q \tag{3.27}$$

where C_I, which is proportional to "electron content" along the signal path, is constant with respect to frequency (but not time) and where f_p and f_q are the current estimates of USO rates for the two frequencies in the given RF band. Based on these expressions, the ionospheric range error in Eq. (3.23) becomes

$$R_I = -c \frac{C_I}{f_p \, f_q} \; . \tag{3.28}$$

Thus, the effective frequency for the range error due to the ionosphere is the geometric mean of the two tone frequencies in the given RF band:

$$\bar{f} = \sqrt{f_p f_q} \; . \tag{3.29}$$

Let $\bar{f}_K$ and $\bar{f}_{Ka}$ denote the effective frequencies for K and Ka band, respectively, computed according to this expression.

Once the ranges, $\hat{R}_K$ and $\hat{R}_{Ka}$, have been obtained for K and Ka band, each at a given time $\tilde{t}_p$, as outlined above, the standard dual-band combination is applied to obtain the ionosphere-corrected range:

$$\hat{R}(\tilde{t}_p) = \frac{\bar{f}^2_{Ka} \, \hat{R}_{Ka}(\tilde{t}_p) - \bar{f}^2_K \, \hat{R}_K(\tilde{t}_p)}{\bar{f}^2_{Ka} - \bar{f}^2_K} \tag{3.30}$$

These range values are generated at a nominal rate of 10 S/s. It is beyond the scope of this report to analyze errors in this calibration technique (arising, for example, from the ionosphere f^{-3} and f^{-4} terms).

The errors in dual-band-calibrated range can be modeled by applying Eq. (3.30) separately to each of the terms in Eq. (3.18), exclusive of the ionosphere term. Nondispersive terms (i.e., terms that are the same for K and Ka bands) pass through Eq. (3.30) without changing form, while dispersive terms (i.e., terms that are different for K and Ka bands) must be appropriately mapped through Eq. (3.30) (e.g., see system noise in Appendix B). In parallel with range processing, a measure of the ionosphere electron content along the signal path can be obtained at a 10-S/s rate by subtracting R_{Ka} and R_K and scaling appropriately.

3.1.5 Digital Filtering to Generate Range, Range Rate and Range Acceleration

As indicated in Fig. 2-2, the ionosphere-corrected range values generated at a nominal rate of 10 S/s are passed through three parallel digital filters to extract the observables for range $(\hat{R})$, range rate $(\hat{\dot{R}})$ and range acceleration $(\hat{\ddot{R}})$ at a selected output rate. This illustration adopts an output rate of 0.1 S/s, which would provide at least 4 samples per harmonic cycle for gravity harmonics, up to approximately degree 135. (Given a 5400-s orbital period, a 0.1-S/s rate yields 0.1 S/s*5400 s/135 = 4 samples/cyc for J135.) In one step, these digital filters reduce the sample rate, apply a low-pass filter with a nominal bandwidth of approximately 0.05 Hz, and, for range rate and acceleration, differentiate to appropriate order, as detailed in Appendix A. To model this step with adequate accuracy for many purposes, one can simply differentiate (if appropriate), reduce the rate to the selected output rate (e.g., 0.1 S/s), and apply a rectangular filter with the selected bandwidth (e.g., 0.05 Hz) to the signal and noise spectra.

Since this stage of processing is carried out on the ground and is relatively fast, it is feasible to pass the same data through this step more than once to obtain a number of output rates, with each tailored to a given user's requirements. For example, a "high" output rate of 0.5 S/s might be required by users investigating more rapid local-gravity variations, while a "low" rate of 1 sample per 30 s might be required by users who plan to estimate only the lower gravity coefficients (e.g., up to 45x45) and wish to limit computation time.

3.2 Frequency Domain Filter Applied to USO Instability

This subsection derives the frequency-domain filter that is effectively applied to USO noise by the dual-1-way operation in Eq. (3.14). It is assumed all samplers and down-converters (K-band and L-band) on a given GRACE satellite are driven by the same oscillator (USO) and that clock offsets and "long-term" drifts (e.g., > 20,000 s) are corrected as outlined above (see case a) on the basis of GPS clock-synchronization measurements. With this approach, the timetags, $\tilde{t}_p$ and $\tilde{t}_q$, are still corrupted by "short-term" (e.g., < 20,000 s) USO phase noise. For this case, the overall error in range due to short-term USO phase noise can be isolated in one expression as follows. First, separate in $\tilde{s}_p$ and $\tilde{s}_q$ the timetag instability due to short-term USO noise at each satellite:

$$\Delta\tilde{s}_p = \frac{\eta_p(t)}{f_p} \tag{3.31}$$

for satellite p and

$$\Delta\tilde{s}_q = \frac{\eta_q(t)}{f_q} \tag{3.32}$$

for satellite q, where η_x is now short-term USO phase noise. When these expressions are substituted in Eq. (3.19) for the USO part of $\tilde{s}_p$ and $\tilde{s}_q$, the range error caused by the short-term USO-noise part of the timetag error becomes

$$\Delta R_{tt} \equiv c\,\frac{f_q - f_p}{f_p + f_q}\left(\frac{\eta_p}{f_p} - \frac{\eta_q}{f_q}\right) \tag{3.33}$$

The overall range error due to short-term noise from both USOs becomes

31

$$\Delta R_{osc} \equiv \Delta R_{tt} + \Delta R_{\phi} \tag{3.34}$$

where ΔR_{ϕ} is given by Eq. (3.21), but with the η_x now containing only short-term USO noise. For USO-p, the resulting range error is approximately given by

$$\Delta R_{osc} \approx \frac{c}{f_p + f_q}\left[\frac{f_q - f_p}{f_p}\,\eta_p(t) + \eta_p(t) - \eta_p(t-\tau)\right] \tag{3.35}$$

which can be rewritten as

$$\Delta R_{osc} \approx \frac{c}{f_p + f_q}\left[\frac{f_q}{f_p}\,\eta_p(t) - \eta_p(t-\tau)\right] \tag{3.36}$$

with a similar result for USO-q. For purposes of error analysis, it is assumed in these expressions that time can be represented by true time and that the delay error terms in Eqs. (3.3) and (3.4) can be ignored under the assumption they are small compared to τ (e.g., $<< \tau \approx 1$ ms).

The frequency response of Eq. (3.36) can be obtained by substituting a unit-amplitude complex input at a given frequency:

$$\eta_p(t) = e^{2\pi i f t} \tag{3.37}$$

in units of cycles. The result is a filter (transfer function), to be denoted by G_o in power, that converts from cycles to µm:

$$\boxed{\sqrt{G_o(f)} \approx \frac{\lambda_e}{2}\left|\left(\frac{f_q}{f_p} - e^{-2\pi i f \tau}\right)\right|} \tag{3.38}$$

where the effective wavelength at RF is defined by

$$\lambda_e \equiv \frac{2c}{f_p + f_q}\,. \tag{3.39}$$

This filter applies to "oscillator-like" noise sources, namely those error sources that simultaneously enter both phase and timetags as described above. If the wavelength is omitted in this expression (i.e., cycles are mapped to cycles), this function predicts the gain experienced by a noise component at frequency f. This "dual-1-way-range" filter is used in Appendix B to propagate "short-term" USO random noise.

If GPS data can be used to correct for both long-term <u>and</u> short-term USO noise, as discussed in case-b above, then the range-error filter for USO noise can be computed by <u>nominally</u> setting $f_p/f_q = 1$ in Eq. (3.38) (but not in Eq. (3.39)). In this approach, timetag errors are not caused by USO instability but by GPS clock-synch errors, and the range error in Eq. (3.19) must be separately computed for the new error source, GPS clock synchronization.

In dual-1-way-range processing, the total error in measured range due to the noise from the USOs is a factor of $\sqrt{2}$ lower than that one would compute for a single USO, after accounting for self cancellation. When the noise for the two USOs (see Eq. (3.21) and (3.32)) is root-sum-squared, the result is a factor of $\sqrt{2}$ increase (assuming noise-equivalent USOs). However, due to the sum of frequencies in the denominator of Eq. (3.14), Eq. (3.38) contains a factor of two in the denominator. The $\sqrt{2}$ increase combined with the factor of two decrease results in an overall decrease of $\sqrt{2}$ relative to the noise one would compute for a single USO, even though two USOs contribute.

For $\tau = 0.667$ ms (i.e., for intersatellite range ≈ 200 km), a plot of the dual-1-way filter in Eq. (3.38) is presented in Fig. 3-2 for Ka band as a function of frequency f and as a function of Δf, where Δf is the RF offset between RF tones (i.e., $f_p = 32$ GHz and $f_q = 32\ \text{GHz} + \Delta f$). For convenience in assessing gain, the plot omits the multiplicative λ factor so that cycles are mapped to cycles. This plot displays the filter that dual-1-way ranging would apply to the noise from one USO and must therefore effectively be used twice in analyzing GRACE observables.

As indicated in Fig. 3-2, the gravity-signal band extends from approximately 0.1 mHz to 18 mHz, for an assumed maximum degree value of 100. Note in Fig. 3-2 that the RF offset Δf causes large changes in filter effectiveness, in that noise reduction is improved by up to two orders of magnitude when Δf is reduced from 10 MHz to 0.1 MHz. Thus, the smallest practical RF offset should be implemented. In this report, the offset is set at a nominal value of 0.5 MHz. With this offset and $\tau = 0.667$ ms, the dual-1-way filter effectively reduces USO noise by a factor of 10^{-5} to 4×10^{-5}. As a result, a USO with stability on the order of 10^{-13} can effectively deliver stability on the order of 10^{-17} to 10^{-18} in measured biased range.

The importance of reducing the RF offset can be alternatively demonstrated as follows. For $\tau = 0.667$ ms and a signal band of 0.1 to 18 mHz, $f\tau$ ranges between 0.07 and 12 μcycle so that one can expand the exponential in Eq. (3.38) to obtain

$$\boxed{\sqrt{G_o(f)} \approx \frac{\lambda_e}{2}\left(\left(\frac{\Delta f}{f_p}\right)^2 + (2\pi f\tau)^2\right)^{1/2}} \,. \tag{3.40}$$

Thus, the $\Delta f/f_p$ term that is caused by the interaction of RF offset (Δf) and sampler timing variations starts to become a lower limit on the filter amplitude (i.e., a limit on the effectiveness of dual-1-way ranging in reducing USO noise) when f falls below, say,

$$f < \frac{1}{2\pi\tau}\frac{|\Delta f|}{f_p} \,. \tag{3.41}$$

For $\Delta f = 500$ kHz, $f_p = 32$ GHz, and $\tau = 0.667$ ms, this cross-over occurs at approximately 4 mHz, which is a relatively high frequency in terms of the gravity-signal band of 0.1 to 18 mHz (up to degree 100). Thus, a lower Δf would be preferable. A Δf value of 100 kHz would cause a crossover at 0.75 mHz, which is well placed near the lower edge of the range-signal band. Even though 100 kHz is a preferred offset, a higher value of closer to 500 kHz will be implemented, as mentioned above.

33

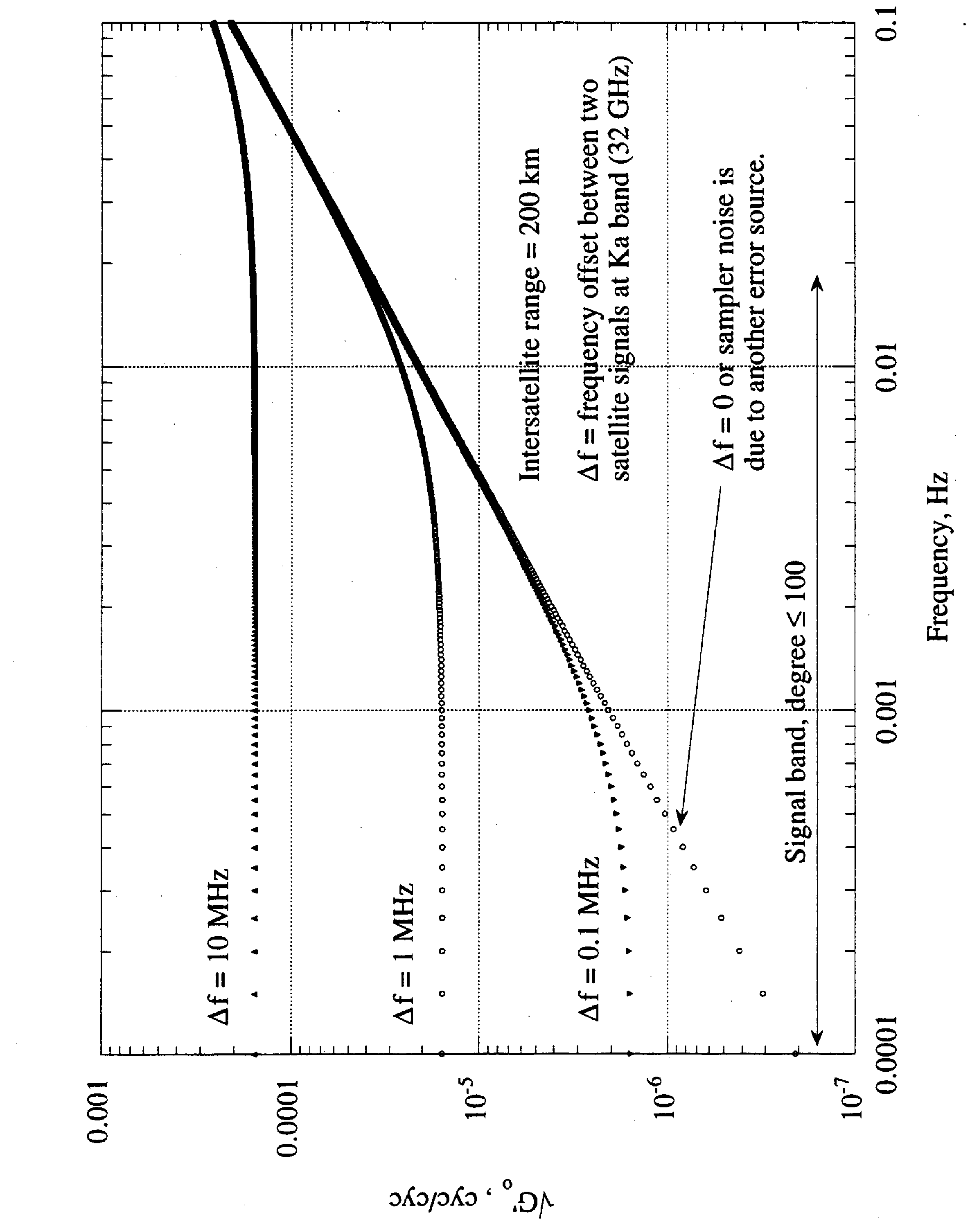

Figure 3-2 Effective Filter Applied to Oscillator Noise by Dual-1-Way Ranging

As mentioned under case-b above, if GPS clock synchronization is sufficiently accurate, short-term as well as long-term variations in timetags can be corrected on the basis of GPS data. Such GPS corrections would be advantageous when GPS-synchronization errors across the signal band (e.g., 0.1 to 18 mHz) are smaller than the corresponding USO errors. When such GPS calibration is justified for both long- and short-term clock synchronization, the frequency response for USO noise is obtained by nominally setting $\Delta f = 0$ in Eq. (3-40), which yields $\sqrt{G_o} \approx \pi f \tau$, in cyc/cyc, as plotted in Fig. 3-2. In this case, an additional range error in the signal band has to be computed to account for GPS short-term clock-synch errors. Such errors are not analyzed in this report. However, if GPS clock synchronization can be used for short-term timetag corrections (case b), note in Fig. 3-2 that the contribution of USO noise at lower signal-band frequencies is substantially reduced (e.g., by approximately a factor of ten at f = 0.36 mHz (2/rev) relative to case-a GPS corrections with $\Delta f = 0.5$ MHz).

3.3 Geometric Component of the Intersatellite Observables

If one neglects all effects except geometric, the range observable, as defined above, is equal to the "light-time" range between satellites as "observed" in a GRACE satellite frame. This subsection derives expressions for the instantaneous geometric component of range and its derivatives in an "inertial" geocentric frame. For purposes of approximate error analysis, one can ignore the important but readily modeled transit-time effect due to satellite motion that appears in light-time solutions for range.

Let $\mathbf{x}_p$ and $\mathbf{x}_q$ be the vectors from Earth's center of gravity (CG) to GRACE satellites p and q, respectively, in the geocentric frame, both at the same observation time. The "range vector" between the satellites is given by

$$\mathbf{R} = \mathbf{x}_q - \mathbf{x}_p \tag{3.42}$$

and the range by

$$R = |\mathbf{x}_q - \mathbf{x}_p| \tag{3.43}$$

which can be rewritten as

$$\boxed{R = \mathbf{R} \cdot \hat{\mathbf{R}}} \tag{3.44}$$

where $\hat{\mathbf{R}}$ is the unit vector along $\mathbf{R}$ defined by

$$\hat{\mathbf{R}} \equiv \frac{\mathbf{R}}{R} \, . \tag{3.45}$$

The square of range can be expressed as

$$R^2 = \mathbf{R} \cdot \mathbf{R} \, . \tag{3.46}$$

Differentiation with respect to time gives

$$2R\dot{R} = 2\,\mathbf{R} \cdot \dot{\mathbf{R}} \tag{3.47}$$

so that the range rate becomes

$$\boxed{\dot{R} = \dot{\mathbf{R}}\cdot\hat{\mathbf{R}}}\ . \tag{3.48}$$

Thus, the geometric component of intersatellite range rate is obtained by projecting the time derivative of the range vector onto the unit vector along the line of sight between satellites.

The geometric component of range acceleration is obtained by differentiating Eq. (3.48):

$$\ddot{R} = \ddot{\mathbf{R}}\cdot\hat{\mathbf{R}} + \dot{\mathbf{R}}\cdot\dot{\hat{\mathbf{R}}}\ . \tag{3.49}$$

Based on Eq. (3.45), the time derivative of the unit vector can be expressed as

$$\dot{\hat{\mathbf{R}}} = \frac{d}{dt}\left(\frac{\mathbf{R}}{R}\right) \tag{3.50}$$

or

$$\dot{\hat{\mathbf{R}}} = \frac{\dot{\mathbf{R}}}{R} - \frac{\dot{R}\hat{\mathbf{R}}}{R}\ . \tag{3.51}$$

Based on this expression and Eq. (3.49), range acceleration becomes

$$\boxed{\ddot{R} = \ddot{\mathbf{R}}\cdot\hat{\mathbf{R}} + \frac{\dot{\mathbf{R}}_\perp\cdot\dot{\mathbf{R}}_\perp}{R}} \tag{3.52}$$

in which

$$\dot{\mathbf{R}}_\perp \equiv \dot{\mathbf{R}} - (\dot{\mathbf{R}}\cdot\hat{\mathbf{R}})\hat{\mathbf{R}} \tag{3.53}$$

is the component of $\dot{\mathbf{R}}$ in the plane perpendicular to $\mathbf{R}$.

The three quantities (range, range rate, and range acceleration), given by Eqs. (3.44), (3.48), and (3.52), respectively, approximately represent the geometric components of the three observables that can be provided by dual-1-way ranging. This report focuses primarily on range acceleration. In acceleration, the second ("centrifugal") term in Eq. (3.52) complicates that observable relative to the other two observables. The centrifugal term is an important term that must be accurately accounted for in full computations of $\ddot{R}$ [Srinivas Bettadpur, private communication, 1997]. For example, for GRACE orbits (nearly coplanar and polar, circular, equiradial), the $\ddot{\mathbf{R}}\cdot\hat{\mathbf{R}}$ term and the centrifugal term can each be approximated as a sum of a relatively large, "constant" component and smaller variations due to orbit differences and gravity harmonics. The constant components of the two terms are of nearly equal magnitude (approximately 0.3 ms^{-2} for $R = 200$ km) and opposite sign so that they cancel one another in the sum forming $\ddot{R}$. With regard to the smaller gravity-harmonic variations, the centrifugal term can make contributions to $\ddot{R}$ that are on the same order as the $\ddot{\mathbf{R}}\cdot\hat{\mathbf{R}}$ term, particularly for high-order harmonics [Willy Bertiger, private communication, 1998]. A crude approach for analyzing the range acceleration observable is presented in Section 4.

3.4 Nongravitational Acceleration

The geometric components of measured range, range rate, and acceleration modeled above contain both gravitational and nongravitational effects. Nongravitational acceleration will be referred to as "drag" for brevity, even though nongravitational effects other than drag are present. The acceleration vector between satellites can be decomposed as

$$\ddot{\mathbf{R}} = \mathbf{a}_q - \mathbf{a}_p + \ddot{\mathbf{R}}_d \tag{3.54}$$

where $\mathbf{a}_x$ is the gravity-induced acceleration vector for satellite x and $\ddot{\mathbf{R}}_d$ is the drag-induced range acceleration, equal to the difference of drag vectors between satellites. Concurrent measurements of the drag vector are to be made on each satellite by a three-axis accelerometer so that drag can be accurately accounted for in modeling. Current plans call for one axis of each accelerometer to be accurately aligned along the line of sight between satellites so that drag along $\hat{\mathbf{R}}$ is nearly directly measured for each satellite. The difference of the output of the two accelerometers for this axis is a measure of differential nongravitational acceleration (i.e., $\ddot{\mathbf{R}}_d \cdot \hat{\mathbf{R}}$) along the line of sight between satellites.

SECTION 4

CRUDE ESTIMATION OF GEOID ERRORS USING THE ACCELERATION OBSERVABLE

4.1 The Approach for Error Analysis

Because of the simplicity allowed by particular, crude but useful approximations, the acceleration observable, $\ddot{R}$, has been selected for error analysis in this report. The hypothetical "fitting" approach assumed here is presented for purposes of crude error analysis and is not intended as a prescription for actual observable fitting. In this approach, only gravity coefficients and instrumental parameters are "estimated." It is assumed the orbits (e.g., x, y, and z for each satellite for each observation) are known with sufficient accuracy, for purposes of error analysis, to compute "direct partials" with respect to gravity coefficients. The "indirect" components of the partials arising from orbit variation, which are considered in Appendix C, are neglected in this error analysis.

The acceleration observable is first corrected on the basis of known orbits to remove the centrifugal term (see Subsection 3.3). As indicated by Eq. (3.52), the geometric component of the corrected acceleration ("projected acceleration") is given by

$$\ddot{R}_p = \ddot{\mathbf{R}} \cdot \hat{\mathbf{R}} \tag{4.1}$$

so that range acceleration can be simply modeled as the projection of the second derivative of the range vector onto the unit vector along the line of sight between satellites.

The acceleration observable is then corrected for drag (see Subsection 3.4). By combining Eqs. (3.54) and (4.1) and subtracting accelerometer-derived drag, $\ddot{\mathbf{R}}_d \cdot \hat{\mathbf{R}}$, one obtains gravity-induced range acceleration:

$$\boxed{\ddot{R}_g = (\mathbf{a}_q - \mathbf{a}_p) \cdot \hat{\mathbf{R}}} \tag{4.2}$$

where $\mathbf{a}_x$ is the gravity-induced acceleration vector for satellite x. The unit vector, $\hat{\mathbf{R}}$, is the actual instantaneous unit vector between satellites, computed in this analysis on the basis of "known" orbits. With this model, an error analysis of acceleration can be carried out relatively simply without the complexity of integration associated with the range and range-rate observables. Since it is assumed that the orbit positions are known, this model for acceleration is linear with respect to gravity coefficients so that calculation of the fit partials is straightforward and only needs to be carried out once, without the iterations sometimes required in simultaneous estimation of orbit and gravity parameters. The "point-mass" (1/r) term can be ignored under the assumption that it can be accurately "removed" based on known orbit positions. The only estimated parameters are the gravity coefficients and parameters to account for instrumental offsets and drifts (e.g., from the accelerometers).

Omission of the centrifugal term can be viewed in another way. Note that omission of the centrifugal term casts the hypothetical acceleration observable in Eq. (4.1) in the same form as range and range rate in Eqs. (3.44) and (3.48), respectively. Given this similarity in form and equivalent observable filtering, some have hypothesized that geoid errors would be on the same order of magnitude when estimated on the basis of $\ddot{\mathbf{R}} \cdot \widehat{\mathbf{R}}$, $\dot{\mathbf{R}} \cdot \widehat{\mathbf{R}}$ or $\mathbf{R} \cdot \widehat{\mathbf{R}}$. It is outside the scope of this report to theoretically quantify the relative accuracy of geoids obtained from these three observable types. However, a comparison of the simulation results obtained here for corrected acceleration with results subsequently obtained from full range-rate simulations [private communication, Srinivas Bettadpur, 1998] indicated that geoid errors due to USO noise are on the same order of magnitude, after approximate corrections to adjust to equivalent orbits, input-noise level, etc.

The acceleration approach adopted in this section, though crude, can provide useful insights and results with regard to the geoid errors caused by a wide range of error sources found in the intersatellite observables. The geoid errors derived here on the basis of this approach are reasonable estimates for errors that would theoretically result from this special, hypothetical "fitting" method.

4.2 Model for Range Acceleration

Based on the preceding subsection, the adopted model for the corrected acceleration observable is given by

$$\ddot{R}_f = (\mathbf{a}_q - \mathbf{a}_p) \cdot \widehat{\mathbf{R}} + a_c \tag{4.3}$$

where $\ddot{R}_f$ is the fit model for acceleration at a given time point (after the corrections discussed above). The term $(\mathbf{a}_q - \mathbf{a}_p) \cdot \widehat{\mathbf{R}}$ is the model for the acceleration due to gravity acting alone along the line-of-sight between satellites, where $\mathbf{a}_x$ is the gravitational acceleration vector for satellite x at the time of the observation. The term a_c represents an instrumentation model (for example, a piecewise constant, linear, or quadratic function of time with a specified update interval). The purpose of the instrumental term is to absorb offsets and slow drifts due to instrumentation, such as those caused by the accelerometers.

4.3 Partial Derivatives of Corrected Acceleration

This subsection derives partial derivatives of corrected acceleration with respect to gravity coefficients, based on Eq. (4.3). Since it is assumed that the positions of the satellites as a function of time in earth-fixed coordinates are known with sufficient accuracy, the arguments (r, θ, λ) of the spherical harmonics presented below can be treated as known quantities in a least squares fit. Similarly, the unit vector between satellites, $\widehat{\mathbf{R}}$, can be computed on the basis of "known" orbits. It is emphasized that the partials computed below on the basis of Eq. (4.3) are specific to this hypothetical model and are developed here for crude error analysis and not for actual parameter estimation.

The Earth's potential (e.g., see [5]) can be represented in terms of spherical harmonics as

$$U(r,\theta,\lambda) = \frac{\mu}{a_e} \sum_{n=0} \left(\frac{a_e}{r}\right)^{n+1} \sum_{m=0}^{n} N_{nm} P_{nm}(\theta) \left[C_{nm} \cos(m\lambda) + S_{nm} \sin(m\lambda)\right] \tag{4.4}$$

where r, θ, λ are the usual spherical coordinates of radius, latitude, and longitude, respectively, with respect to the Earth's center of gravity; $a_e = 6378.1363$ km is the Earth's semimajor axis;

$\mu = 398600.4415 \text{ km}^3/\text{s}^2$ is GM; n, m are the degree and order of the spherical harmonic; P_{nm} is the unnormalized associated Legendre polynomial; and C_{nm}, S_{nm} are the dimensionless gravity coefficients to be estimated. N_{nm} is Kaula's normalization factor [5], which is computed as $[(n-m)!\,(2n+1)(2-\delta_{m0})/(n+m)!]^{1/2}$.

The gravitational acceleration for satellite p at point $(r_p, \theta_p, \lambda_p)$ is given by the gradient of the potential:

$$\mathbf{a}_p = \frac{\partial U}{\partial r_p}\hat{\mathbf{u}}_r + \frac{1}{r_p}\frac{\partial U}{\partial \theta_p}\hat{\mathbf{u}}_\theta + \frac{1}{r_p\cos\theta_p}\frac{\partial U}{\partial \lambda_p}\hat{\mathbf{u}}_\lambda \tag{4.5}$$

where$(\hat{\mathbf{u}}_r, \hat{\mathbf{u}}_\theta, \hat{\mathbf{u}}_\lambda)$ are point-specific unit vectors in the direction of increasing $(r_p, \theta_p, \lambda_p)$, respectively.

Since $\ddot{R}_f$ in Eq. (4.3) is linear with respect to the gravity coefficients, one can easily compute the partial of $\ddot{R}_f$ with respect to the gravity coefficients, C_{nm} and S_{nm}, for given satellite orbital positions using the preceding equations:

$$\frac{\partial \ddot{R}_f}{\partial C_{nmj}} = \left(\tilde{\mathbf{a}}_{nmj,q} - \tilde{\mathbf{a}}_{nmj,p}\right) \cdot \hat{\mathbf{R}} \tag{4.6}$$

where the C_{nm}, S_{nm} coefficients are now represented by one symbol, C_{nmj}, for which j denotes the $\cos(m\lambda)$ $(j = 1)$ or $\sin(m\lambda)$ $(j = 2)$ term. (The zero-order terms have only the cosine term.) The coefficient/satellite-specific accelerations, $\tilde{\mathbf{a}}_{nmj,x}$, can be obtained for satellite x by combining Eqs. (4.4) and (4.5) with $C_{nmj} = 1$:

$$\tilde{\mathbf{a}}_{nmj,x} = \frac{\partial \tilde{U}_{nmj,x}}{\partial r_x}\hat{\mathbf{u}}_r + \frac{1}{r_x}\frac{\partial \tilde{U}_{nmj,x}}{\partial \theta_x}\hat{\mathbf{u}}_\theta + \frac{1}{r_x\cos\theta_x}\frac{\partial \tilde{U}_{nmj,x}}{\partial \lambda_x}\hat{\mathbf{u}}_\lambda \tag{4.7}$$

where

$$\tilde{U}_{nmj,x} \equiv \frac{\mu}{a_e}\left(\frac{a_e}{r_x}\right)^{n+1} N_{nm}P_{nm}(\theta_x)\, T_j(m\lambda_x) \tag{4.8}$$

in which $T_j(m\lambda)$ is equal to $\cos(m\lambda)$ for $j = 1$ and $\sin(m\lambda)$ for $j = 2$. Since $\ddot{R}_f$ is linear with respect to the coefficients (under the assumption the orbits are known), only a one-time computation of the partials is necessary to allow subsequent parameter estimation. These partials will be referred to as the "sensitivity partials."

4.4 Examples of Corrected-Acceleration Partials

Examples of sensitivity-partial magnitudes, based on the preceding approximate expressions and defined as root-mean-squared (RMS) values over time, are presented in Figs. 4-1 through 4-7 as a function of coefficient degree and order for selected orbits. The total sensitivity partial is plotted as well as the three components (r, θ, λ) in Eq. (4.7) that contribute to the total partial. Note that the units of the partials are $\mu\text{m s}^{-2}\text{ cmG}^{-1}$, indicating that the units of the gravity coefficients have been changed from dimensionless to "cmG", or cm of equivalent geoid error, by multiplying each gravity coefficient by a_e, the Earth's semimajor axis. Partial plots are very useful in estimating and understanding overall sensitivity of projected acceleration to gravity coefficients, relative sensitivity with regard to components, and the expected improvement to be gained as longitude separation is increased from 0 to 200 km. It is emphasized that these partials assume that the non-negligible centrifugal term in the acceleration observable can be "removed" on the basis of known orbits (see Section 3).

These RMS partials have been computed from Eq. (4.6) on the basis of perfect circular polar orbits, with both satellites at the same altitude (300 km) and period (5430 s). The adopted orbital period (5430 s) provides a good ratio ($\approx$111 orbits / 7 days) between orbital period and sidereal day (86164.1 s), given an adopted fit interval of 7 days and estimated harmonics up to (50, 50). That is, up to the highest order, 50, this period leads to a good sampling distribution in longitude. In reality, the two orbits can not be perfectly circular as a result of variations due to gravity harmonics. Such effects can be ignored since such orbit variations are relatively small and do not significantly change partials for the purposes of this error analysis. A similar argument can be made with respect to matched circular orbits and eccentricity.

The first three plots, Figs 4-1, 4-2, and 4-3, present RMS partials as a function of order for degree 50, based on three longitude separations: 0, 40 and 200 km, respectively. The orbits provide a latitude separation of 200 km between satellites and an altitude of 300 km for both satellites. In Fig. 4-3 for 200-km longitude separation, the radial component of the partials has very small magnitude relative to the other components. Thus, it is the latitude and longitude components that will provide strength to the multiparameter solution for these gravity coefficients. Furthermore, the coefficients with $m = n$ have total partials that are approximately a factor of two weaker than the coefficient with $m = 0$, and therefore would have coefficient errors approximately twice as large, given a flat noise spectrum and minimal correlation between estimated parameters. Note that the longitude component provides the strength at large m values (as a result of the 200-km longitude offset) while the latitude component provides the strength at smaller m values.

Fig. 4-3 is to be contrasted with Fig. 4-1 for which the longitude separation has been reduced to zero. As one would expect, the longitude component of the partials for zero longitude separation is zero, greatly weakening sensitivity at large m values. For example, at $m = n = 50$, the total partial is smaller by approximately a factor of 20 relative to the 200-km-longitude-separation case in Fig. 4-3. As shown in Section 6, the solution for gravity coefficients with $m = n$ shows a considerable degradation for the 0-km-separation case relative to the 200-km-separation case, reflecting this loss in partial magnitude. Note in Fig. 4-1 that the partial for $m = n = 50$ is 50 times smaller than the partial for $m = 0$, indicating the zonals (harmonics with $m = 0$) will be determined with far greater accuracy than the sectorials (harmonics with $m = n$). Further, note that the radial component for $m = n = 50$ makes a large contribution to the total partial, approximately half, leading to a value of approximately 0.0028 rather than 0.0015 $\mu\text{m s}^{-2}\text{ cmG}^{-1}$. For lower-order partials, the radial component makes a smaller contribution, rapidly decreasing in importance as order decreases. For example, the radial component increases the partial from 0.0033 to 0.0047 $\mu\text{m s}^{-2}\text{ cmG}^{-1}$ for $m = 49$ and from 0.0052 to 0.0067 $\mu\text{m s}^{-2}\text{ cmG}^{-1}$ for $m = 48$.

Fig. 4-2, which presents degree-50 partials for a 40-km separation, shows that even a small longitude separation can substantially strengthen the partials for the $m = n$ coefficients: the RMS partial for $m = n$ is improved by approximately a factor of two for the 40-km separation relative to 0-km separation.

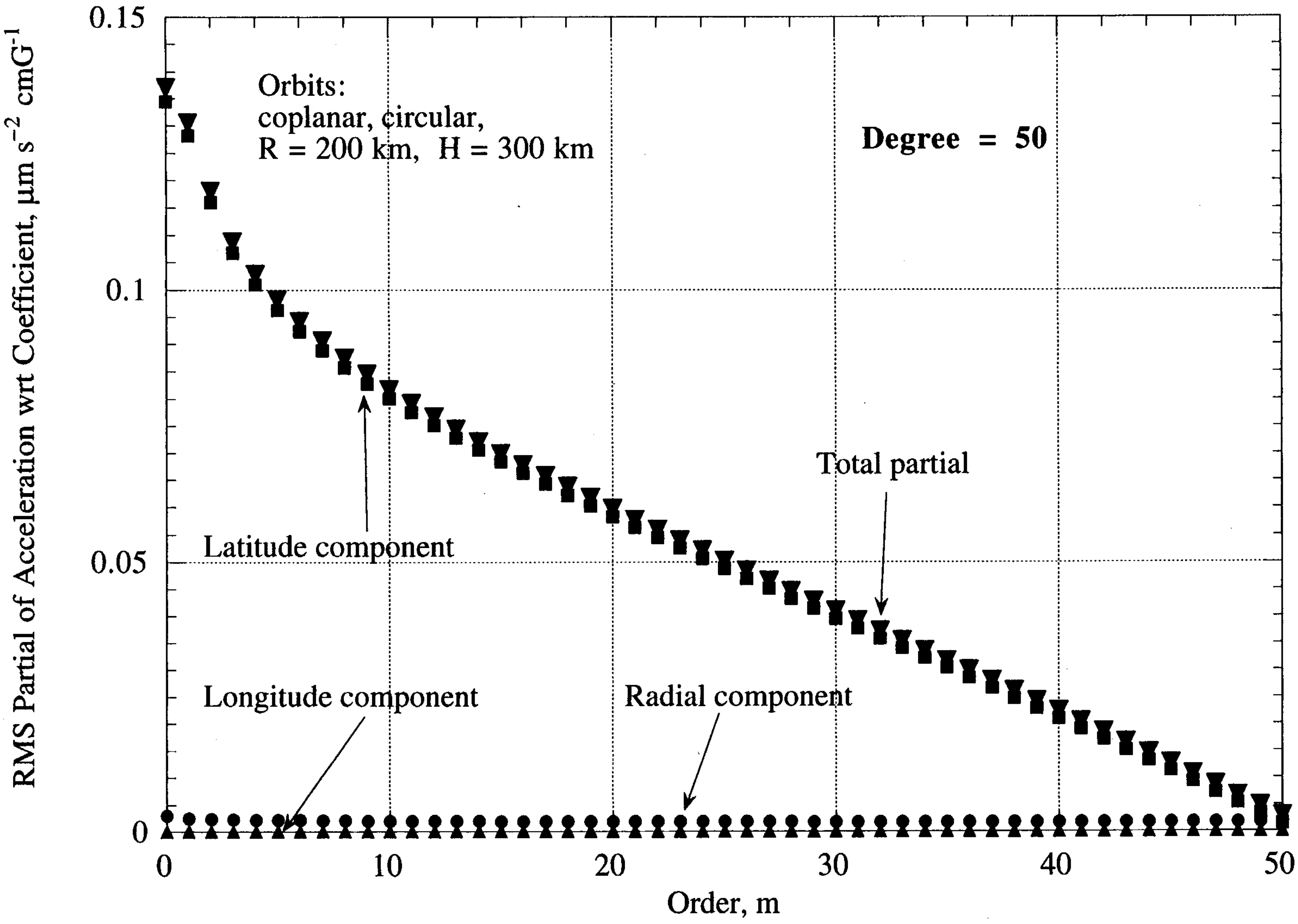

Figure 4-1 Example A of RMS Partials of Corrected Acceleration Observable with respect to Degree-50 Coefficients

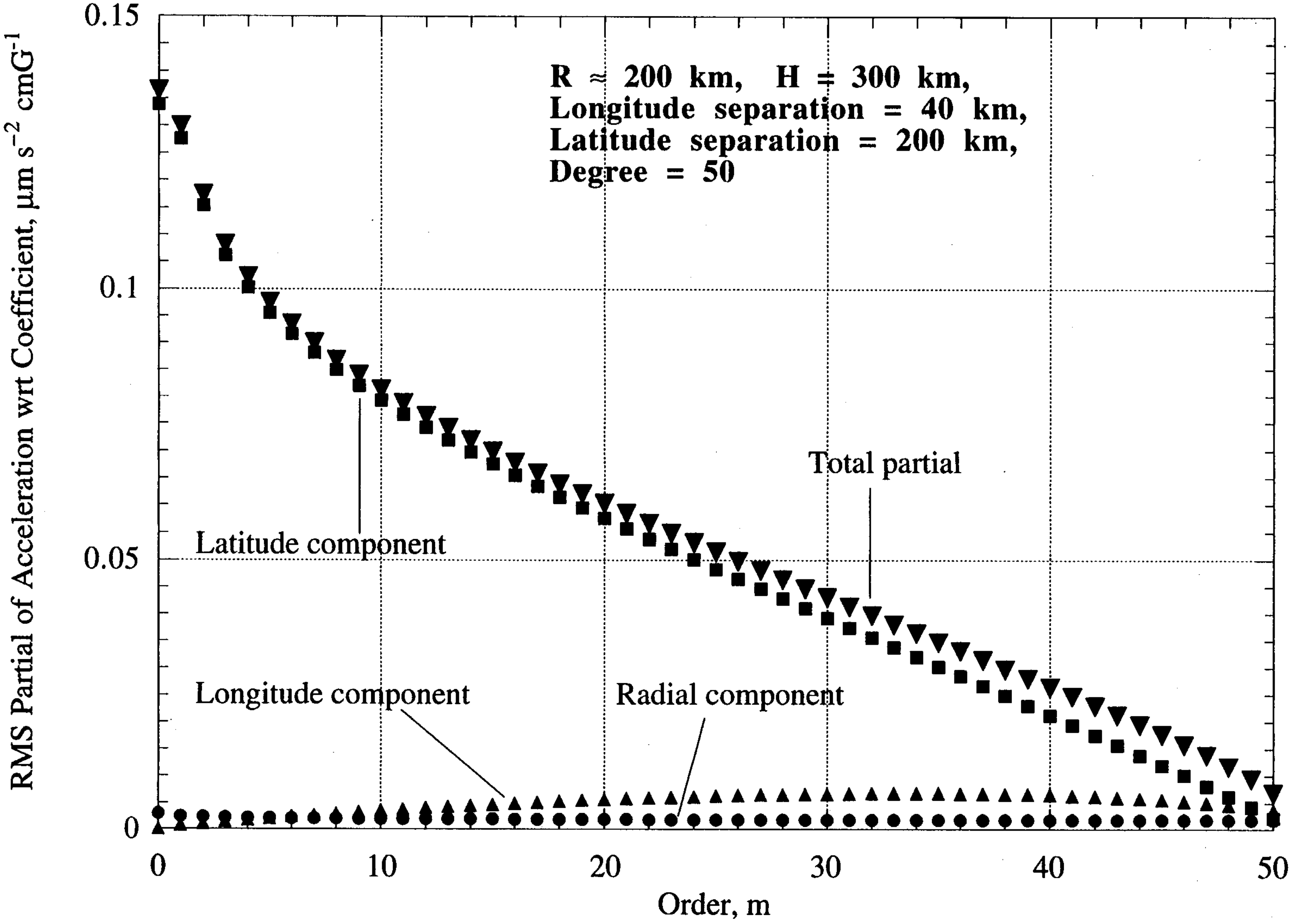

Figure 4-2 Example B of RMS Partials of Corrected Acceleration Observable with respect to Degree-50 Coefficients

HU

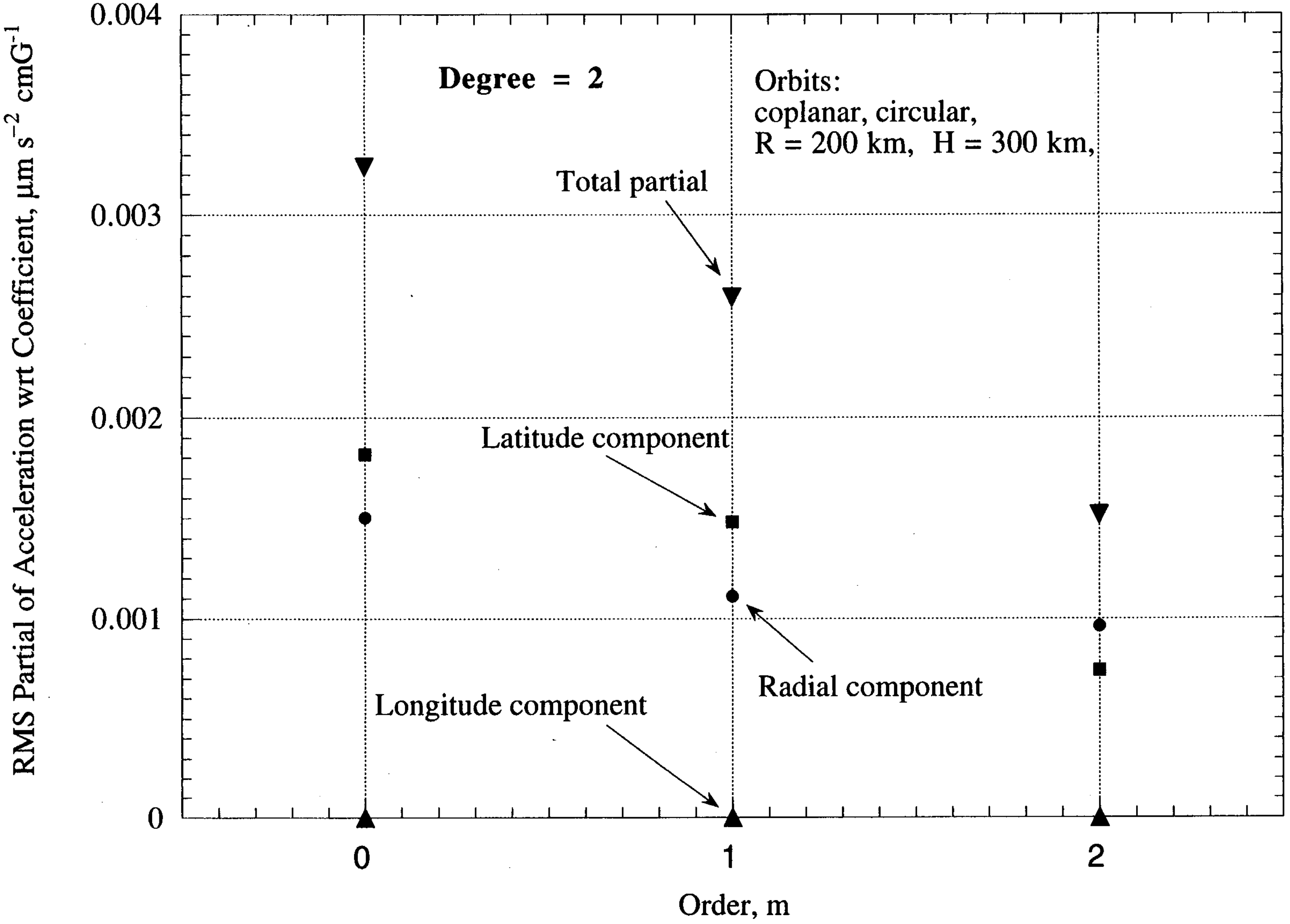

Figure 4-4 Example A of RMS Partials of Corrected Acceleration Observable with respect to Degree-2 Coefficients

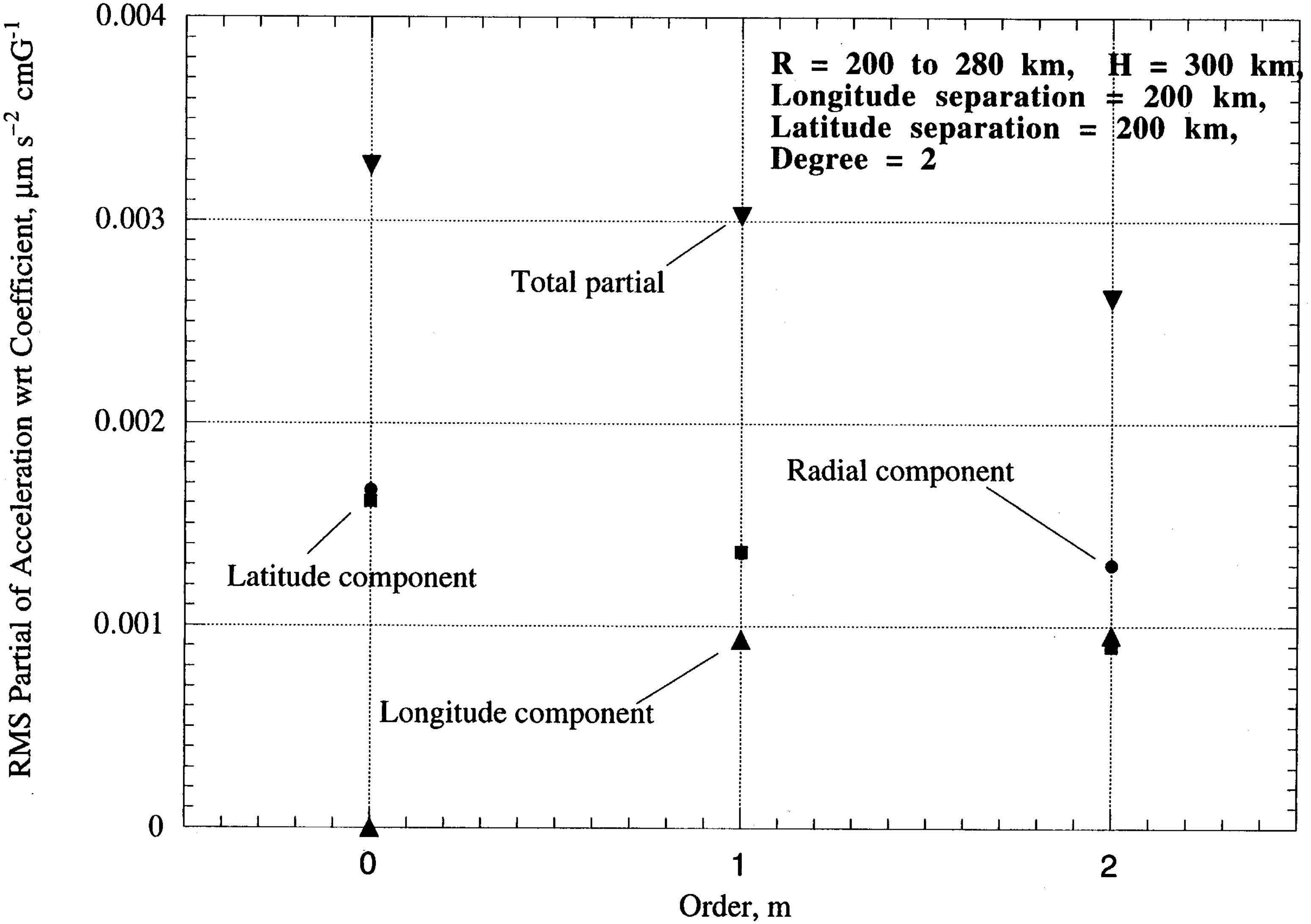

Figure 4-5 Example B of RMS Partials of Corrected Acceleration Observable with respect to Degree-2 Coefficients

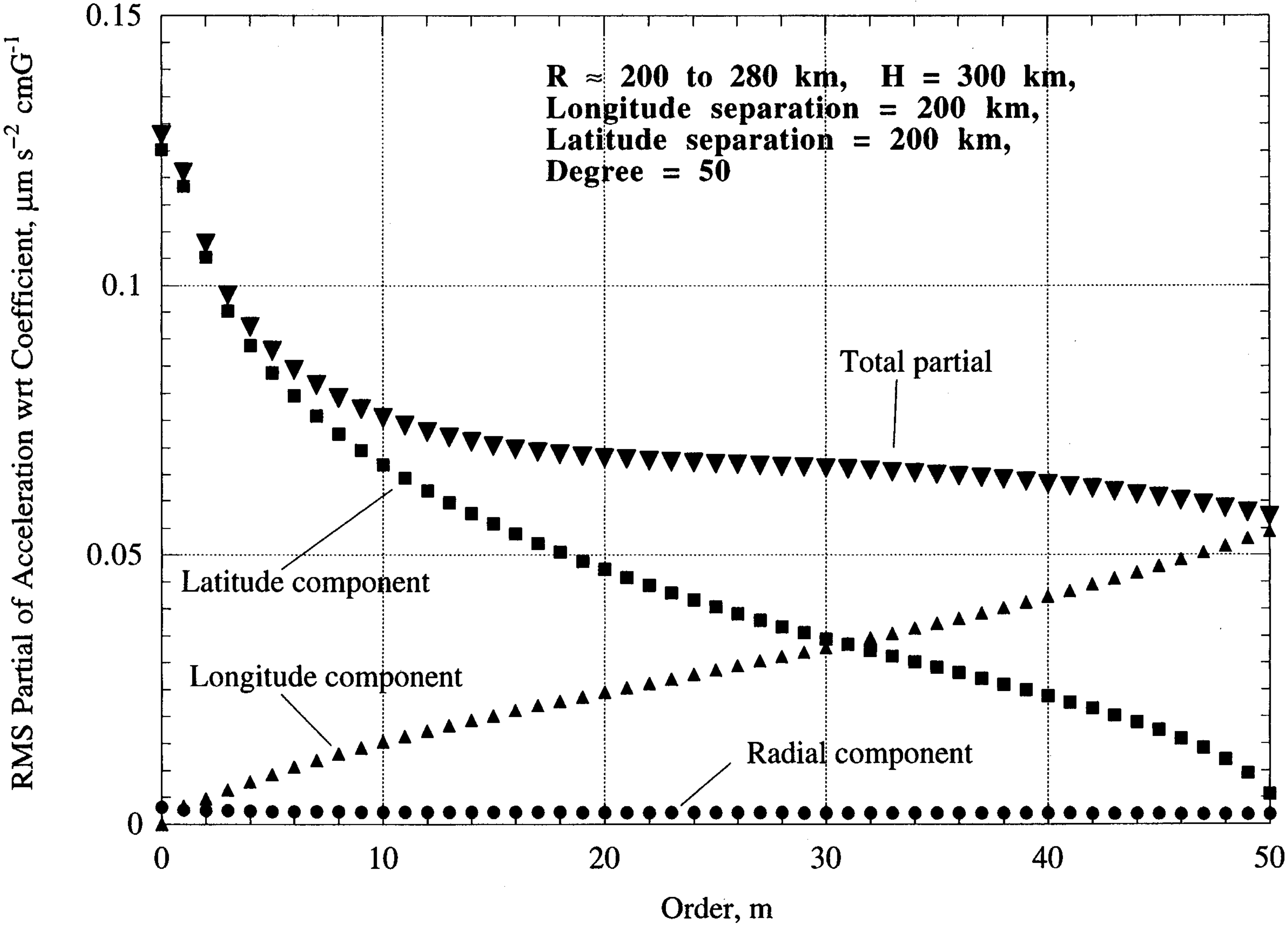

Figure 4-3 Example C of RMS Partials of Corrected Acceleration Observable with respect to Degree-50 Coefficients

Similar relative behavior is displayed by the partials for other degree values, although, as degree decreases, the difference between zonals and sectorials is less striking for zero-longitude separation and the improvement provided by increasing longitude separation is less pronounced. For comparison and reference, the corresponding RMS partials for degree 2 are presented as a function of order in Figs. 4-4 and 4-5 for longitude separations of 0 and 200 km, respectively. For zero-longitude separation in Fig. 4-4, the m = n = 2 partial is approximately half the m = 0 partial (approximately 0.0015 vs. 0.0032 in $\mu m\ s^{-2}\ cmG^{-1}$). Increasing longitude separation from 0 to 200 km improves the (2, 2) partial from 0.0015 to 0.0026 $\mu m\ s^{-2}\ cmG^{-1}$.

RMS partials as a function of degree for selected orders are presented in Figs. 4-6 and 4-7 for longitude separations of 0 and 200 km, respectively. For the 200-km longitude separation in Fig. 4-7, partial magnitude increases, from approximately 0.0015 to the vicinity of 0.1 in units of $\mu m\ s^{-2}\ cmG^{-1}$, as degree increases from 2 to 40, and then decreases. Partials for m = n are smaller than the m = 0 partials, by up to a factor of three in this plot. Partials for other values of order fall between these two curves, increasing in magnitude as order decreases.

For the 0-km longitude separation in Fig. 4-6, the zonal partials are approximately the same as for 200-km separation. Here, however, sectorial partials are far smaller than zonal partials, by as much as a factor of 70 for degree 70. Partial magnitude increases rapidly as m decreases near n, for fixed degree. For example, for degree 70, partial magnitude is approximately a factor of 1.7 larger at m = n – 1 than at m = n and a factor of three larger at m = n – 3 than at m = n.

These partial plots further emphasize the advantage provided by larger longitude separation. However, because of disadvantages such as the increased attitude/pointing demands, increased drag, and increased sensitivity to accelerometer-weak-axis noise, a 200-km separation does not appear to be advisable. If these considerations constrain the orbits to a longitude separation of 0 to 40 km, the differences in partials outlined above lead to a large imbalance between errors for low-order and high-order gravity coefficients for a given degree value.

45

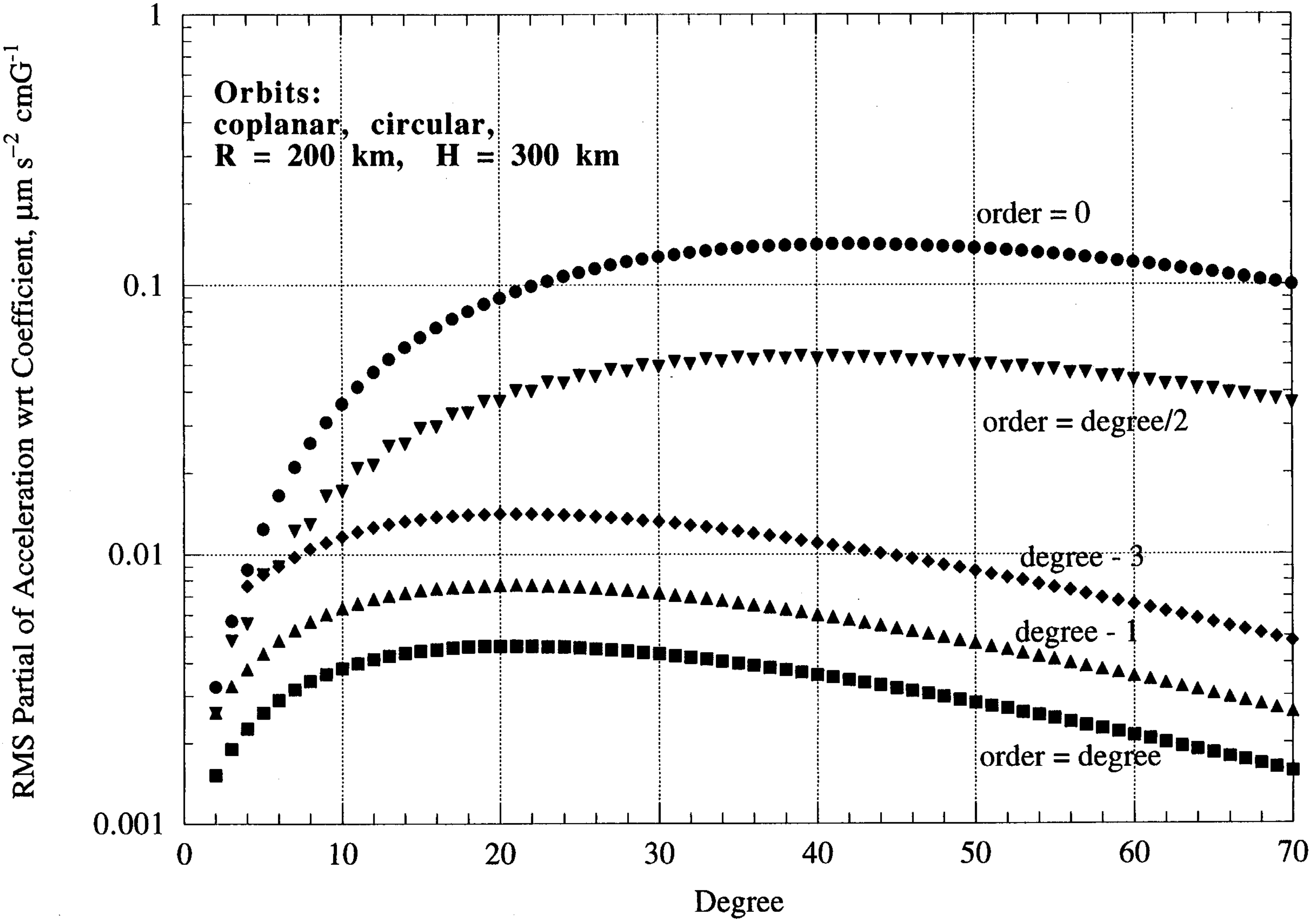

Figure 4-6 Example of RMS Partials of Corrected Acceleration Observable with respect to Gravity Coefficient for Coplanar Satellites

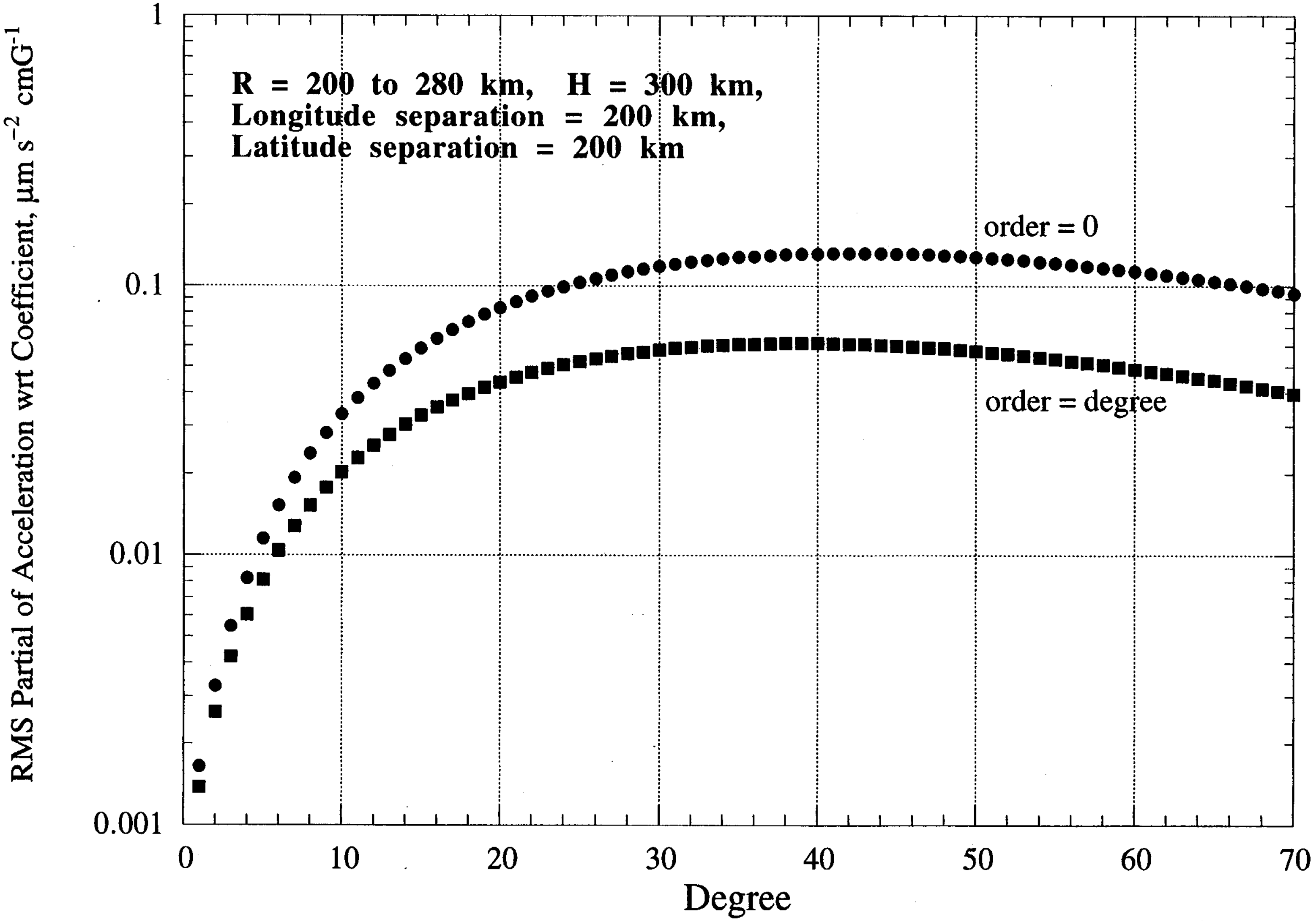

Figure 4-7 Example of RMS Partials of Corrected Acceleration Observable with respect to Gravity Coefficient, Given 200-km Longitude Separation

SECTION 5

APPROXIMATE CLOSED-FORM EXPRESSIONS FOR GEOID ERRORS

Based on an approximate, zero-correlation approach to error propagation, this section derives simple closed-form expressions for gravity-coefficient errors, when estimates are based on the hypothetical fits to the acceleration observable. As in the rest of this report, the gravity-estimation approach assumes orbit positions are known and ignores the centrifugal term. In spite of approximations, the approach can provide insight into gravity estimation and fast, useful, order-of-magnitude estimates of gravity coefficient errors. The closed-form approximations require only a few seconds to run on a PC compared to a day for a full simulation on more powerful computers.

5.1 The Zero-Correlation Approximation

If correlations between estimated parameters are ignored, an approximate closed-form approach can be developed to propagate errors in acceleration from each error source through the fit equation to obtain associated geoid errors. Appendix C analyzes modeling of instrumental drifts (i.e., $f^{-\alpha}$ noise) and suggests such errors (i.e., from modeling $1/\sqrt{f}$ accelerometer noise in the case of the corrected acceleration observable) do not degrade the error approximations of this section. Appendix C also presents a simple example that suggests that orbit parameters "separate" from gravity coefficients in a gravity solution. Appendix D analyzes the importance of correlations between estimated gravity coefficients and suggests that those correlations do not excessively degrade the error approximations of this section. Other approximations in the approach, as discussed below, can cause additional errors.

The fit model for the acceleration observables is the approximate expression given by Eq. (4.3), with the same estimated parameters assumed in that section. In this section, error-propagation equations are derived by approximately propagating errors through the standard least-squares fit equation:

$$\hat{\mathbf{p}} = \left(\mathbf{A}^{T}\mathbf{A}\right)^{-1}\mathbf{A}^{T}\mathbf{y} \tag{5.1}$$

where $\hat{\mathbf{p}}$ is the vector of estimated parameters, given by

$$\hat{\mathbf{p}} = \left(\hat{p}_1, \hat{p}_2, \ .....\hat{p}_{N_p}\right)^{T} \tag{5.2}$$

and **y** is the acceleration observable vector, defined by

$$\mathbf{y} \equiv \left(\ddot{R}_1, \ddot{R}_2, \ ....\ddot{R}_{N_o}\right)^{T} \tag{5.3}$$

where N_p is the number of parameters to be estimated and N_o is the number of observations. **A** is the partials matrix, defined in terms of the partial derivatives of the observable model with respect to the estimated parameters (see Section 4):

$$A_{ij} \equiv \frac{\partial \ddot{R}_i}{\partial p_j} \ , \ i = 1 \text{ to } N_o \ , \ j = 1 \text{ to } N_p \tag{5.4}$$

As in Section 4, the estimated parameters account for an instrumental function, a_c, and the gravity coefficients. Note that Eq. (5.1) represents a "white-noise fit" in that it is "optimal" for observables with noise uncorrelated point to point, with constant standard deviation. This analysis does not consider more complicated fitting approaches.

In fits with Eq. (5.1) for many gravity coefficients (e.g., 100x100), there will be correlations between estimated parameters since the off-diagonal elements in $(\mathbf{A}^T\mathbf{A})^{-1}$ will be nonzero (see Appendix D). These correlations complicate solutions using Eq. (5.1) and impede insight and closed-form analysis. When the correlations are not extreme, useful approximations and intuitive insight can be gained by assuming the correlations are zero, which, in effect, is equivalent to solving for each parameter separately using the one-parameter version of Eq. (5.1). One can easily show that a one-parameter solution for a parameter p is given by

$$\hat{p} = \frac{\sum_{i=1}^{N_o} \frac{\partial y_i}{\partial p} y_i}{\sum_{i=1}^{N_o} \left(\frac{\partial y_i}{\partial p}\right)^2} \tag{5.5}$$

where $y_i = \ddot{R}_i$ for the acceleration observable. Thus, the one-parameter least-squares estimate is obtained forming the dot product of the observation vector with the partials vector and dividing by the square of the length of the partials vector. In other words, the parameter estimate is equal to the component of the observation vector lying along the partials vector, divided by the length of the partials vector.

5.2 Example of a Two-Tone Partial

When satellite position (r, θ, λ) as function of time is substituted in Eqs. (4.5) through (4.8), the spatial dependence of the spherical harmonics leads to a time dependence. Since spherical harmonics have spatial periodicity as a function of latitude and longitude (see Appendix E), the acceleration placed on a satellite by a given harmonic becomes a time function that is a sum of tones whose frequencies are determined by spatial periodicity of that harmonic and the orbital period. For example, for a polar orbit, a degree-50 zonal would lead to a time function with tones at 50/rev, 48/rev, etc., where rev denotes the orbital period.

Since each spherical harmonic leads to a sum of tones in the acceleration observable (see Subsection 5.5 below), it is instructive to determine the nature of a fit like Eq. (5.5) for a simple example. Let the observable be a sum of only two pure sinusoids, each with known phase (orbit is accurately known), known constant frequencies (f_1 and f_2 from harmonic periodicity) and known "subamplitudes" (a_1 and a_2 from harmonic components), but with a common unknown overall amplitude A (the gravity coefficient, C_{nm} or S_{nm}, to be estimated). For this simple example, the i^{th} measured acceleration $\ddot{R}_i$ is given by

$$\ddot{R}_i = A[a_1 \cos(2\pi f_1 t_i) + a_2 \cos(2\pi f_2 t_i)] + \eta \tag{5.6}$$

where η is additive noise, $t_i = i\Delta t$ is the time of observation i, and Δt is the time between observations. The partials become

$$\frac{\partial \ddot{R}_i}{\partial A} = a_1 \cos(2\pi f_1 t_i) + a_2 \cos(2\pi f_2 t_i) \tag{5.7}$$

Based on Eqs. (5.6) and (5.7), Eq. (5.5) becomes

$$\hat{A} \approx \frac{2}{N_o\left(a_1^2 + a_2^2\right)} \sum_{i=1}^{N_o} \left[a_1 \cos(2\pi f_1 t_i) + a_2 \cos(2\pi f_2 t_i)\right] \ddot{R}_i \tag{5.8}$$

where it has been assumed, without significant loss of validity, that the sum extends over exactly an integer number of cycles of f_1 and f_2 so that the mean square partial is given by

$$\bar{P}_A^2 \equiv \frac{1}{N_o} \sum_{i=1}^{N_o} \left(\frac{\partial \ddot{R}_i}{\partial A}\right)^2 = \frac{1}{2}\left[a_1^2 + a_2^2\right] \tag{5.9}$$

Note that the solution for A in Eq. (5.8) essentially extracts the Fourier components of $\ddot{R}_i$ at the two frequencies and combines them according to subamplitude weights. In spite of its simplicity, Eq. (5.8) can be used to illustrate several important features relevant to gravity estimation.

If the signal part of the observable (i.e., $A[a_1 \cos(2\pi f_1 t_i) + a_2 \cos(2\pi f_2 t_i)]$) is substituted in Eq. (5.8) for $\ddot{R}_i$, one can easily show that the estimated amplitude is equal to the actual amplitude (i.e., $\hat{A} = A$) as one would expect. That is, each component in Eq. (5.8) "seeks out" its corresponding component in $\ddot{R}_i$ and contributes an appropriate amount to the estimate so that the correct total amplitude is obtained. On the other hand, if a unit-amplitude noise component at frequency f',

$$\eta_i = \cos(2\pi f' t_i) \tag{5.10}$$

is substituted in place of $\ddot{R}_i$ in Eq. (5.8), one obtains the associated frequency response (in "voltage") of the fit filter for A, which can be approximated for a long fit interval by

$$\sqrt{G_A(f')} \approx \frac{1}{a_1^2 + a_2^2}\left[a_1 \frac{\sin[\pi(f' - f_1)T_{fit}]}{\pi(f' - f_1)T_{fit}} + a_2 \frac{\sin[\pi(f' - f_2)T_{fit}]}{\pi(f' - f_2)T_{fit}}\right] \tag{5.11}$$

where $T_{fit} = N_o \Delta t$ is the fit interval. (This expression, which is for positive f', ignores the smaller "negative-frequency" terms in the fit filter.) Thus, the frequency response of the fit is approximately a sum of two sinx/x terms centered at the two tone frequencies of the signal, each with a width of $1/T_{fit}$ at first null. (See Appendix C for illustrations of various fit filters.) Since the width of each sinx/x is extremely small (e.g., approximately 1 μHz for a nominal fit interval of approximately 13 days), the frequency response of the fit filter is essentially a sum of two extremely sharp spikes whose relative amplitudes are determined by the subamplitudes in the partial derivative. Absolute amplitude for each spike is obtained by dividing subamplitude by the sum of the squares of the subamplitudes.

Let $S_\eta(f)$ (in units of $\mu m^2 s^{-4} Hz^{-1}$ for the acceleration observable) be the spectral noise density at frequency f that describes the noise η in the observable. The error (power) spectrum for A is computed as the product of the (power) fit filter and the spectral noise density: $G_A(f)\, S_\eta(f)$. The variance of total error on A, σ_A^2, is computed by integrating this product over the frequency:

$$\sigma_A^2 = \int_{-\infty}^{\infty} G_A(f)\, S_\eta(f)\, df \tag{5.12}$$

It should be noted that, in actual fits, the frequencies of partial "tones" are not constant as assumed in the derivation of Eq. (5.11) but vary slowly as a result of orbit decay. This variation causes the fit-filter peaks in actual fits to be "smeared out." For error analysis, this change in tone frequency can be ignored and tones can be assigned a fixed "average" frequency. This procedure is equivalent to assigning the orbits a "fixed radius" and computing the gravity-coefficient errors for that radius. This approximation is not expected to significantly degrade error estimates for gravity coefficients at a given (average) altitude. Thus, even though such slow variation is important in actual fits to estimate gravity coefficients, it can be ignored for purposes of approximate error analysis, and effective peak width can be approximated by $1/T_{fit}$.

5.3 Approximate Closed-Form Expression for Gravity-Coefficient Errors

The results of the last subsection can be used to derive a simple closed-form expression for estimating gravity-coefficient errors based on the zero-correlation approximation.

Since one can easily show that the effective bandwidth of a $[\sin(\pi\Delta f T_{fit})/(\pi\Delta f T_{fit})]^2$ peak is $1/T_{fit}$, the noise power (in units of $\mu m^2\ s^{-4}$) contributed by such a peak is $S_\eta(f)/\ T_{fit}$, in the flat-spectrum approximation. (See Appendix C for a heuristic analysis of sidelobes and colored noise spectra.) Thus, based on the two spikes in Eq. (5.11) and associated weights (a_1 and a_2), the error variance on estimated amplitude in Eq. (5.12) can be approximated by

$$\sigma_A^2 \approx 2\frac{\left[a_1^2 S_\eta(f_1) + a_2^2 S_\eta(f_2)\right]}{\left[a_1^2 + a_2^2\right]^2 T_{fit}} \tag{5.13}$$

This approximation to the integral over frequency assumes the spectral noise density is sufficiently well-behaved to allow the $(\sin x/x)^2$ functions to dominate the shape of the integrand (i.e., sidelobe power drops off sufficiently rapidly in the product, $S_\eta(f)\ (\sin x/x)^2$) so that f_1 and f_2 become the respective effective frequencies and $1/T_{fit}$ is a good approximation of effective bandwidth for each $\sin x/x$. The factor of two accounts for fit-filter terms at negative frequencies where it is assumed S_η is single sideband noise that is symmetric about zero frequency. (That is, negative-frequency noise is picked up by the $f' + f_1$ and $f' + f_2$ terms omitted in Eq. (5.11)).

If one tone is dominant (e.g., $a_1 >> a_2$) and if the noise spectrum does not counteract this dominance, Eq. (5.13) becomes

$$\sigma_A^2 \approx 2\frac{S_\eta(f_1)}{a_1^2 T_{fit}} \tag{5.14}$$

which can be rewritten as

$$\boxed{\sigma_A^2 \approx \frac{S_\eta(f_1)}{\overline{P}_A^2\ T_{fit}}} \tag{5.15}$$

where the RMS partial obtained from Eq. (5.9) has been approximated by

$$\overline{P}_A \equiv \frac{1}{\sqrt{2}} a_1 \tag{5.16}$$

The simple form presented in Eq. (5.15) turns out to be a very quick and useful way of computing approximate coefficient errors, as shown later in this section. Within its approximations, it indicates that the error in a given gravity coefficient can be estimated by dividing propagated random noise by RMS partial, where propagated noise is equal to the single-sideband spectral noise found at the effective frequency characterizing the harmonic, within the fit bandwidth (e.g., $\sqrt{S_\eta(f_1)/T_{fit}}$).

If the observable is corrupted by an error tone, Eq. (5.11) indicates that the error tone maps to the estimated gravity coefficient in proportion to the fit filter amplitude found at the frequency of that error tone. Thus, if the error-tone is separated in frequency from all partial tones by an amount substantially greater than $1/T_{fit}$, the error in A due to the error tone will be greatly attenuated. Since $1/T_{fit}$ is on the order of 1 μHz for a 13 day fit, an error tone must be very close to a partial tone in frequency to avoid extreme attenuation. In terms of phase coherence, the error-tone phase must match the phase of a partial tone to substantially better than a cycle to avoid large attenuation. For crude estimates, error-tone attenuation due to phase mismatch can be calculated as $\sin(\pi\Delta\phi)/(\pi\Delta\phi)$, where $\Delta\phi$ is the phase deviation in cycles between error tone and partial tone over the fit interval due to a constant frequency mismatch. If a tone is sufficiently phase coherent with a partial tone for a given gravity coefficient, the coefficient error caused by that tone does not decrease as a function of fit-interval length, as a random error or a phase incoherent tone does. For example, an exactly coherent 2/rev error tone in the acceleration observable, with an amplitude of 5×10^{-5} μm s^{-2}, would map approximately to a geoid error of 0.01 cmG for J2, given the RMS partial of 0.0033 μm s^{-2} cmG^{-1} in Fig. 4-4.

5.4 Approximate Closed-Form Expressions for Root-Mean-Square Partials

This subsection derives approximate closed-form expressions for the RMS partial derivatives of corrected acceleration, Eq. (4.2), with respect to gravity coefficients. The resulting expressions can be used in Eq. (5.15) to approximate errors in gravity coefficients. The derivation assumes the two satellites have circular coplanar orbits with the same radius, the same period and a selected latitude offset. A number of approximations must be made to obtain the final closed-form expression for partials, with the understanding that crude estimates of gravity-coefficient errors are the ultimate goal.

As shown in Section 4, the contribution of the radial component to the partial derivative of acceleration with respect to gravity coefficient is small except for a few values for degree and order (e.g., for m = n or for very low-degree values). For coplanar satellites ($\lambda_1 = \lambda_2$), the contribution of the longitude component is zero. Thus, for this orbit configuration, the latitude component is typically dominant and represents a good general approximation for the total partial. With these assumptions, consideration of Eqs. (4.6), (4.7) and (4.8) indicates that the acceleration partial with respect to gravity coefficient can be approximated by

$$\frac{\partial \ddot{R}}{\partial \tilde{C}_{nmj}} \approx \frac{\mu}{a_e^3} \left(\frac{a_e}{r}\right)^{n+2} N_{nm} \frac{d^2 P_{nm}(\theta)}{d\theta^2} T_j(m\lambda)\, \Delta\theta \tag{5.17}$$

where $\Delta\theta$ is satellite separation in latitude in radians and $T_j(m\lambda)$ represents either $\cos(m\lambda)$ or $\sin(m\lambda)$, depending on the value of j. This approximation sets the radial and longitude components in Eq. (4.5) to zero and assumes the line-of-sight unit vector is approximately equal to the unit vector in the latitude direction (i.e., $\hat{\mathbf{R}} \cdot \hat{\theta} \approx 1$) given the expected small separation in latitude (e.g., a latitude separation of 200-km/6650-km ≈ 0.03 in radians). Further, the difference in accelerations in Eq. (4.6) has been converted to a derivative based on the small angular separation in latitude. The radii of the two satellites are assumed to be equal ($r_1 = r_2 = r$), which implicitly assumes circular orbits. (In reality, the radii are not equal and the orbits are not perfectly circular as a result of non-zero eccentricity and variations due to gravity harmonics. Such effects can be ignored if simple crude error analysis is the goal.) As indicated in Sections 3 and 4, the centrifugal term is neglected in this approximation. The tilde on $\tilde{C}_{nmj}$ denotes a conversion of units for the gravity coefficients from dimensionless to distance units (e.g., by multiplying C_{nmj} by a_e to obtain units of cmG, where the "G" stands for "of geoid"). Thus, the partial has units such as $\mu\text{m s}^{-2}\text{ cmG}^{-1}$.

To illustrate the nature of Legendre polynomials (LPs), Appendix E presents plots for a number of LPs and their derivatives. The first derivative shows up in the gradient of the gravity field in the latitude direction, as in the satellite acceleration in Eq. (4.5). To first approximation (at least when the range between satellites is sufficiently small), the second derivative shows up in the difference in accelerations between satellites, as in Eq. (5.17). As suggested in Appendix E, high degree zonals (m = 0) have strength over the whole globe, but with considerable prominence at the poles and no longitude dependence. The high-degree sectorials (m = n), on the other hand, have their strength concentrated in a narrow band at the equator (approximately $|\text{lat}| < 1.2/\sqrt{n}$ rad half maximum) and are modulated in the longitude direction by $\cos(n\lambda)$.

The RMS value for the acceleration partial is defined by

$$\overline{P}_a(n, m) \equiv \left[\left\langle\left[\frac{\partial \ddot{R}}{\partial \tilde{C}_{nmj}}\right]^2\right\rangle\right]^{1/2} \tag{5.18}$$

where the average is computed for the latitude and longitude values sampled during the fit interval (e.g., 13 days). It will be assumed that the orbits have been selected so that these values of latitude and longitude provide an adequate sampling distribution in latitude and longitude. (That is, the sampling should be close to uniform and provide at least one sample for each 0.25 cycle of harmonic phase in both the latitude and longitude directions. For example, given observables spaced at 10 s for a polar orbit with a period on the order of 5400 s, there are approximately 5400-s/10-s = 540 points per revolution, which provides a sample spacing in the latitude direction of 100/540 = 0.185 cyc for the 100 cyc/rev found in the (100, 0) harmonic. In the longitude direction, given a 13-day fit interval and an orbit period that is sufficiently incommensurate with the sidereal period, there are approximately 400 separate equatorial crossings, which can provide a sample spacing in longitude of approximately 100/400 = 0.25 cyc for the 100 cyc per 2π in longitude found in the (100, 100) harmonic. Under the assumption of adequate angular sampling, the average over observations (time) can be converted, to good approximation, to a continuous average over latitude and longitude.

As suggested by Eq. (5.17), computation of an RMS partial involves averaging the square of the product of two functions: the second derivative of an associated Legendre polynomial times a longitude sinusoid (T_j). Given the assumptions of the preceding paragraph, one can show that the average can be approximately separated into longitude and latitude factors:

$$\left\langle \left[\frac{d^2P_{nm}(\theta)}{d\theta^2} T_j(m\lambda)\right]^2 \right\rangle \approx \left\langle \left[\frac{d^2P_{nm}(\theta)}{d\theta^2}\right]^2 \right\rangle_\theta \left\langle T_j^2(m\lambda)\right\rangle_\lambda \tag{5.19}$$

where the averages are now continuous averages (i.e., integrals over $\theta = -\pi/2$ to $+\pi/2$ and $\lambda = 0$ to 2π). Since T_j is equal to 1.0 or to a sine or cosine function, the longitude average becomes

$$\left\langle T_j^2(m\lambda)\right\rangle_\lambda = 1.0 \quad \text{for } m = 0, j = 1 \tag{5.20}$$

and

$$\left\langle T_j^2(m\lambda)\right\rangle_\lambda \approx 1/2 \quad \text{for } m \neq 0 \tag{5.21}$$

Numerical computations have been used to determine the latitude average of the second derivative of the Legendre polynomial. Based on standard normalization for Legendre polynomials, Figs. 5-1 through 5-5 plot the RMS derivatives for selected values of order (i.e., $m = 0$, $n/2$, $n-2$, $n-1$ and n) as a function of degree from 2 to 100. Fig. 5-6 summarizes those plots for the second derivative. As indicated in the figures (at least for these values of m), the RMS second derivative of a Legendre polynomial can be represented to good approximation by the simple form $\beta\, n^\gamma$ where β and γ depend on m. For example, the RMS second derivative is approximately equal to (n^2, $0.34n^2$, $1.8n$, $1.1n$, $n/2$) for $m = (0, n/2, n-2, n-1, n)$, respectively.

Based on Eqs. (5.17), (5.18), (5.19), (5.29), and (5.21) and the results from Fig. 5-6, the RMS partials of range acceleration with respect to gravity coefficient can be approximated by the simple closed-form expression

$$\boxed{\overline{P}_a(n,\, m) \approx \frac{\mu}{a_e^3}\left(\frac{a_e}{r}\right)^{n+2} F_{nm}\, \Delta\theta} \tag{5.22}$$

where

$$F_{nm} \approx (\sqrt{2}\, n^2,\ 0.48\, n^2,\ 2.5\, n,\ 1.6\, n,\ 0.71\, n) \tag{5.23}$$

for $m = (0, n/2, n-2, n-1, n)$, respectively. In this equation, the standard normalization for associated Legendre polynomials (see Fig. 5-6) has been converted to the Kaula normalization as in Eq. (4.4) by multiplying the standard normalization factor by $\sqrt{2}$ for $m = 0$ and by 2 for $m \neq 0$.

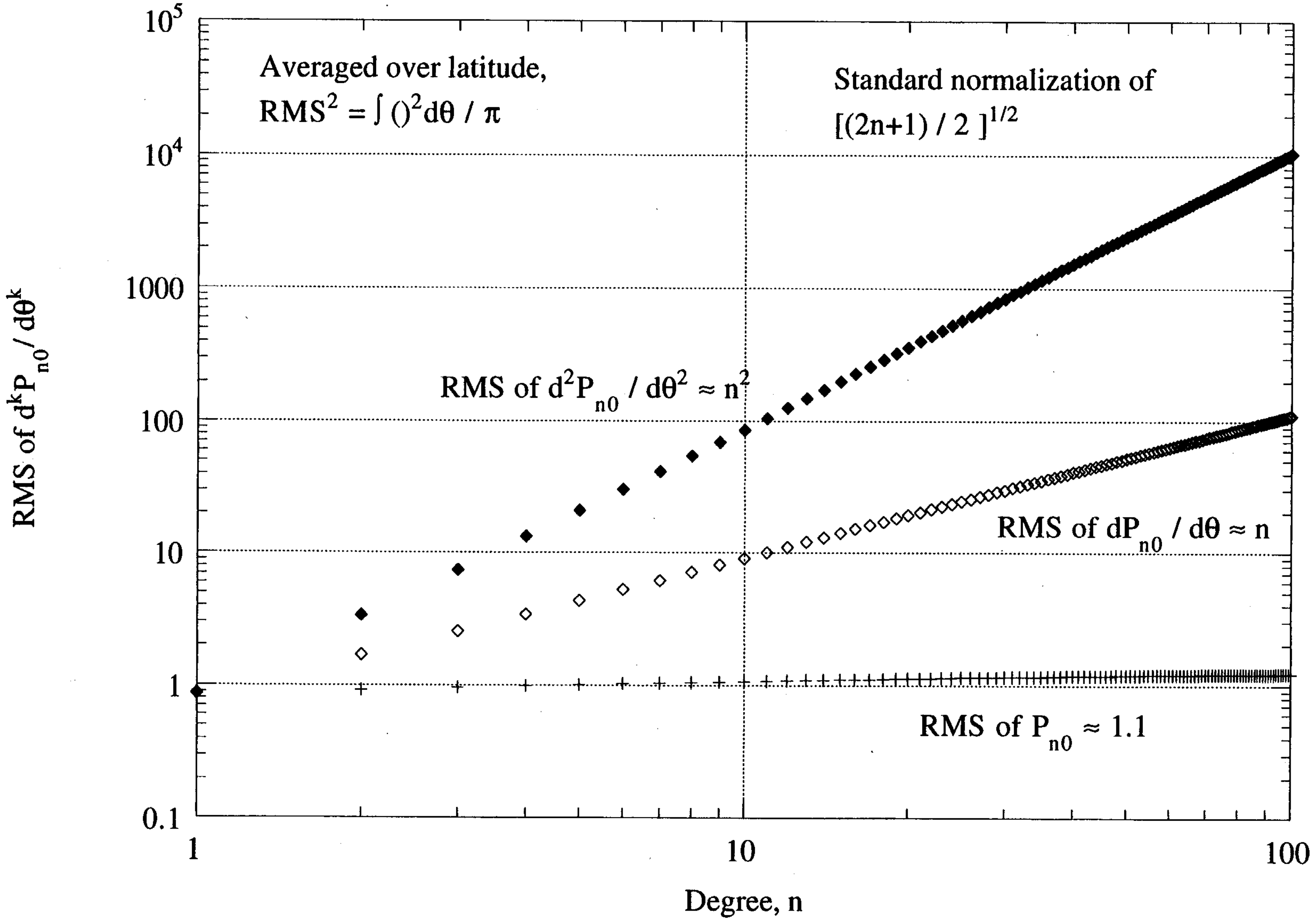

Figure 5-1 RMS Derivatives of Legendre Polynomials with respect to Latitude

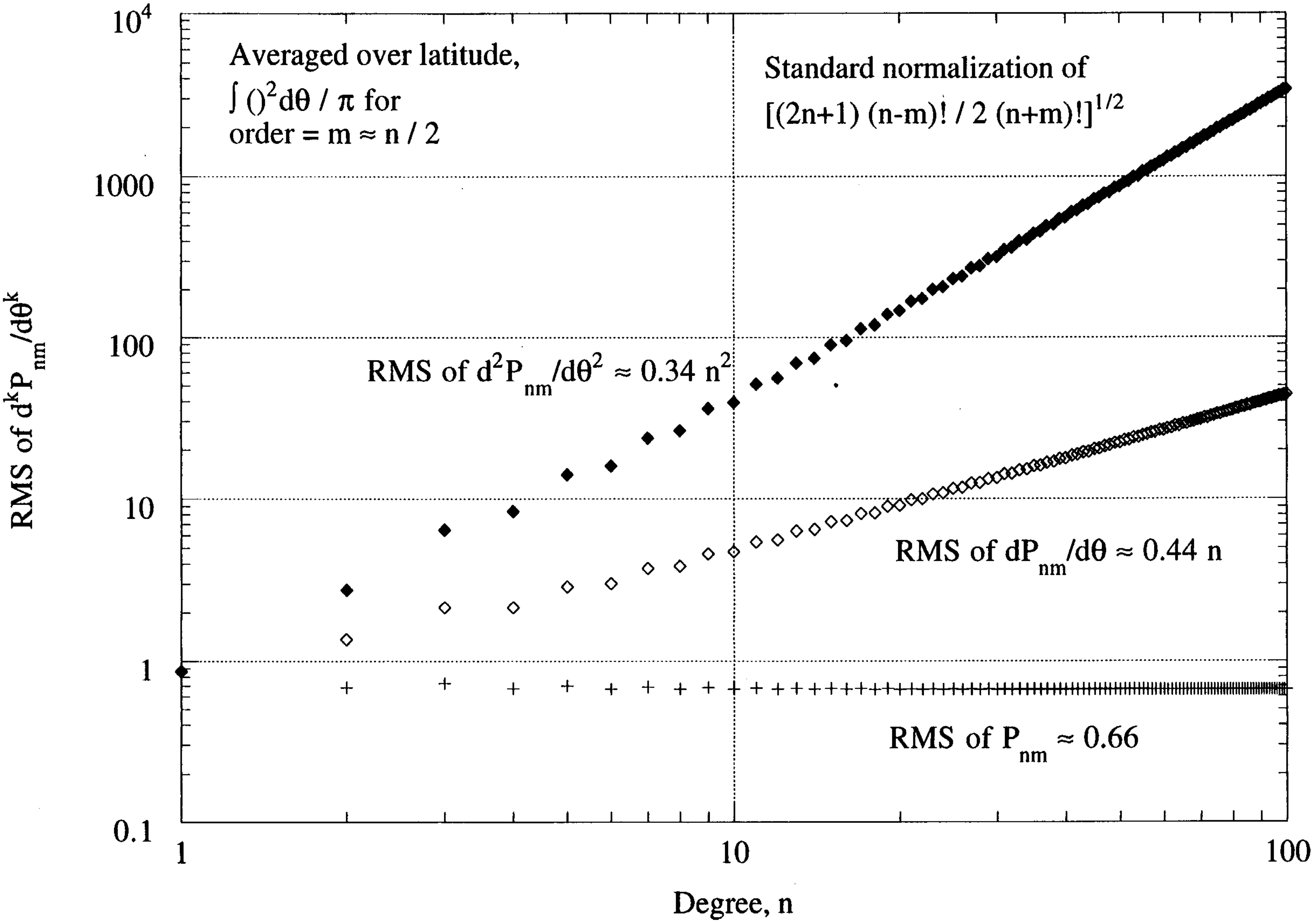

Figure 5-2 RMS Derivatives of Associated Legendre Polynomials with respect to Latitude for Order ≈ Degree/2

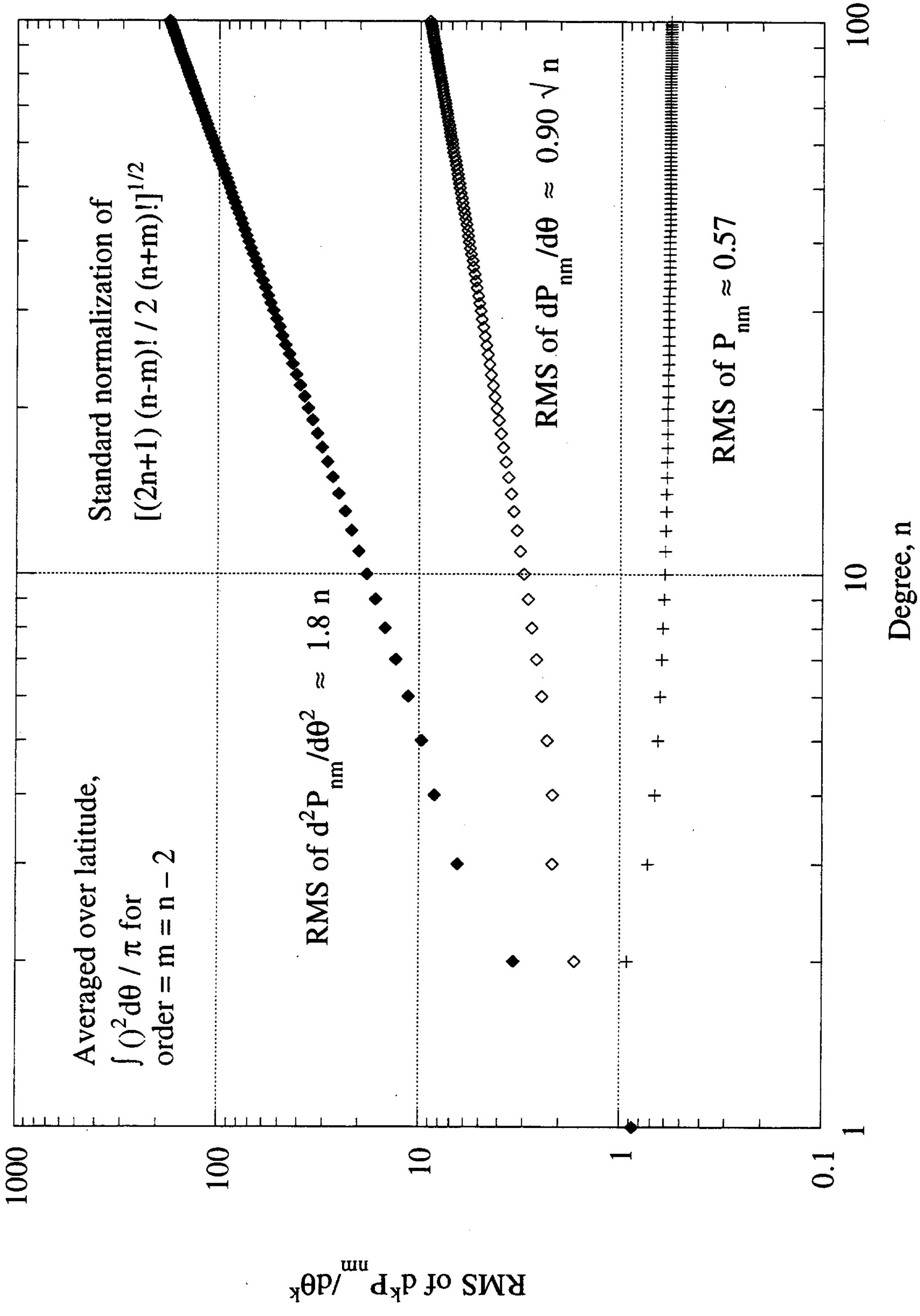

Figure 5-3 RMS Derivatives of Associated Legendre Polynomials with respect to Latitude for Order = Degree − 2

60

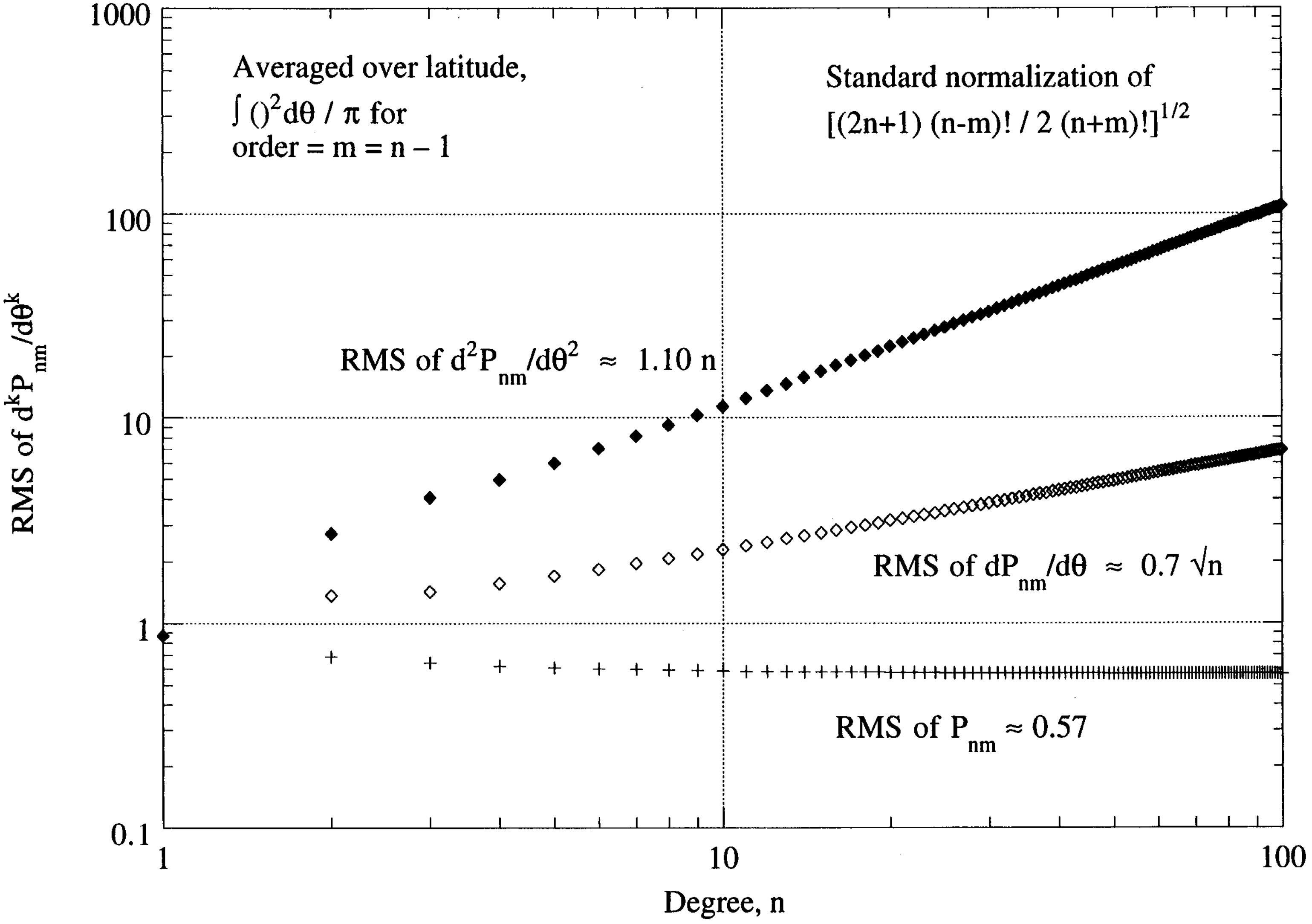

Figure 5-4 RMS Derivatives of Associated Legendre Polynomials with respect to Latitude for Order = Degree – 1

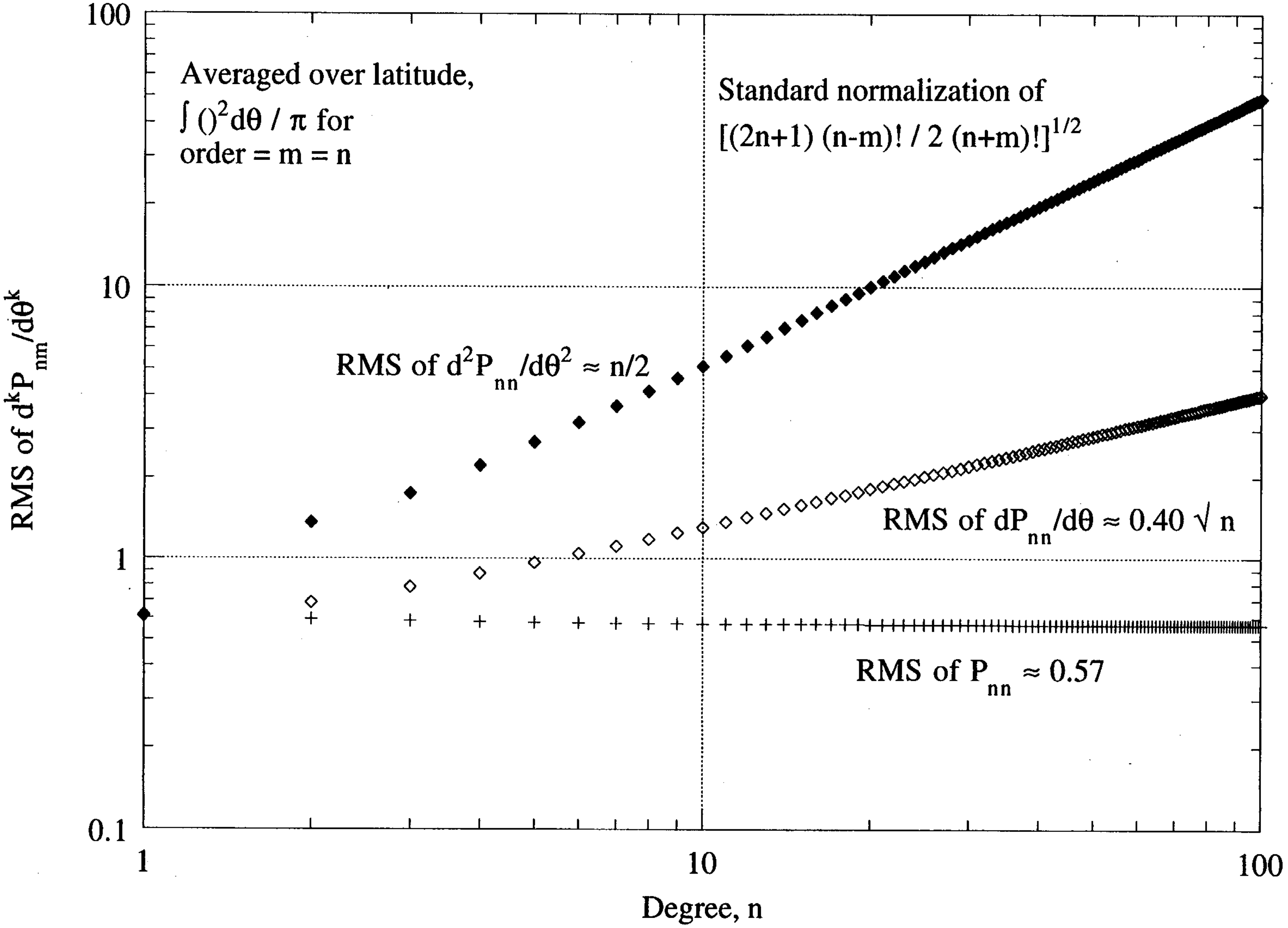

Figure 5-5 RMS Derivatives of Associated Legendre Polynomials with respect to Latitude for Order = Degree

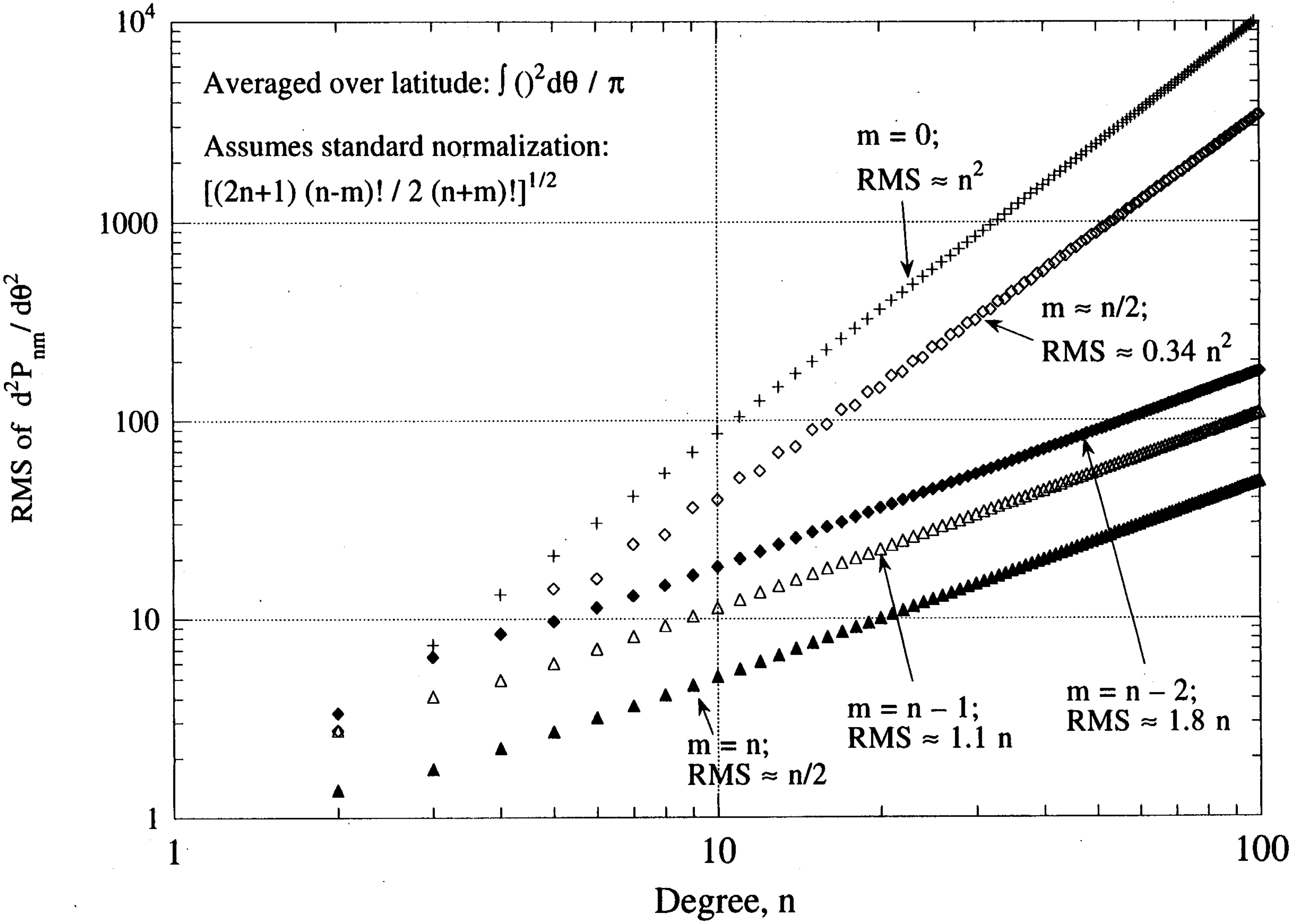

Figure 5-6 RMS of Second Derivative of Legendre Polynomials

Since Eq. (5.22) approximates the latitude difference between satellites through use of a derivative, it can overestimate partial magnitude in the cases of large n, small m for which $\Delta\theta$ is not necessarily small. For the example of n = 100, m = 0, R = 200 km, and H = 300 km, $\Delta\theta$ spans approximately 0.5 cycle of phase in the Legendre polynomial (i.e., approximately 200 km / 400 km where 400 km is the approximate spatial equivalent of 1 cyc in the latitude direction for the (100, 0) harmonic). For this particular case (which is the worst case for these experiment parameters), partial magnitude is overestimated by approximately a factor of 1.5 with respect to this effect. In the spirit of the other crude approximations in this derivation, this effect can be ignored.

Fig. 5-7 plots the approximate expression for RMS partials in Eq. (5.22) as a function of degree (2 to 100) for the selected values of order. These approximate partials are to be compared to the more accurate numerical partials plotted in Fig. 4-6. Figs 4-6 and 5-7 are in crude agreement, with agreement rapidly improving as order decreases. The worst order is m = n, which is approximately a factor of two lower than the numerical result. This difference is due to the radial term, which is neglected in Eq. (5.22) and Fig. 5-7, but is accounted for in the numerical calculation, as plotted separately in Fig. 4-1 and discussed in Section 4. For most values of degree and order, Eq. (5.22) provides a fairly accurate approximation to the partials for coplanar orbits.

The weakness of the latitude component of the partials for high order is a consequence of the weaker latitude dependence of the high-order associated Legendre polynomials. For m = n, the Legendre polynomials have no zero crossings while, for m = 0, there are 2n zero crossings per revolution. The (n, 0) polynomial has a dominant tone at a frequency of n cyc/rev but the (n, n) polynomial has dominant tones at the lowest frequencies of 1/rev and 2/rev. This weaker variability for (n, n) causes the smaller second derivative with respect to latitude indicated by Eq. (5.23). (One can show that the (n, n) polynomial varies in proportion to $[\cos(\theta)]^n$ as a function of latitude. For large n, this function has the largest magnitude at zero latitude and falls off rapidly and monotonically as latitude departs from zero. As indicated by Eq. (4.4), the full spherical harmonic modulates this behavior in the longitude direction with a cosine or sine function covering n cycles in phase per 2π in longitude.)

5.5 Effective Tone Frequencies for Acceleration Partials

The preceding subsection provides a simple closed-form expression for approximating the RMS partial of corrected acceleration with respect to gravity coefficient. Before Eq. (5.22) can be used to estimate coefficient error, however, a method is needed for determining an approximate value for effective tone frequency {i.e., f_1 in Eq. (5.15)} for each partial. Effective tone frequency can be estimated by examining the spectral content of the acceleration partial for each spherical harmonic, as suggested by the fit filter analysis in Appendix C. Examples of discrete Fourier transforms (DFTs) of such partials are presented in Figs. 5-8 through 5-15. These plots assume the orbits are circular and coplanar with the same altitude (300 km) and period (5400 s) and an intersatellite range of 200 km.

63

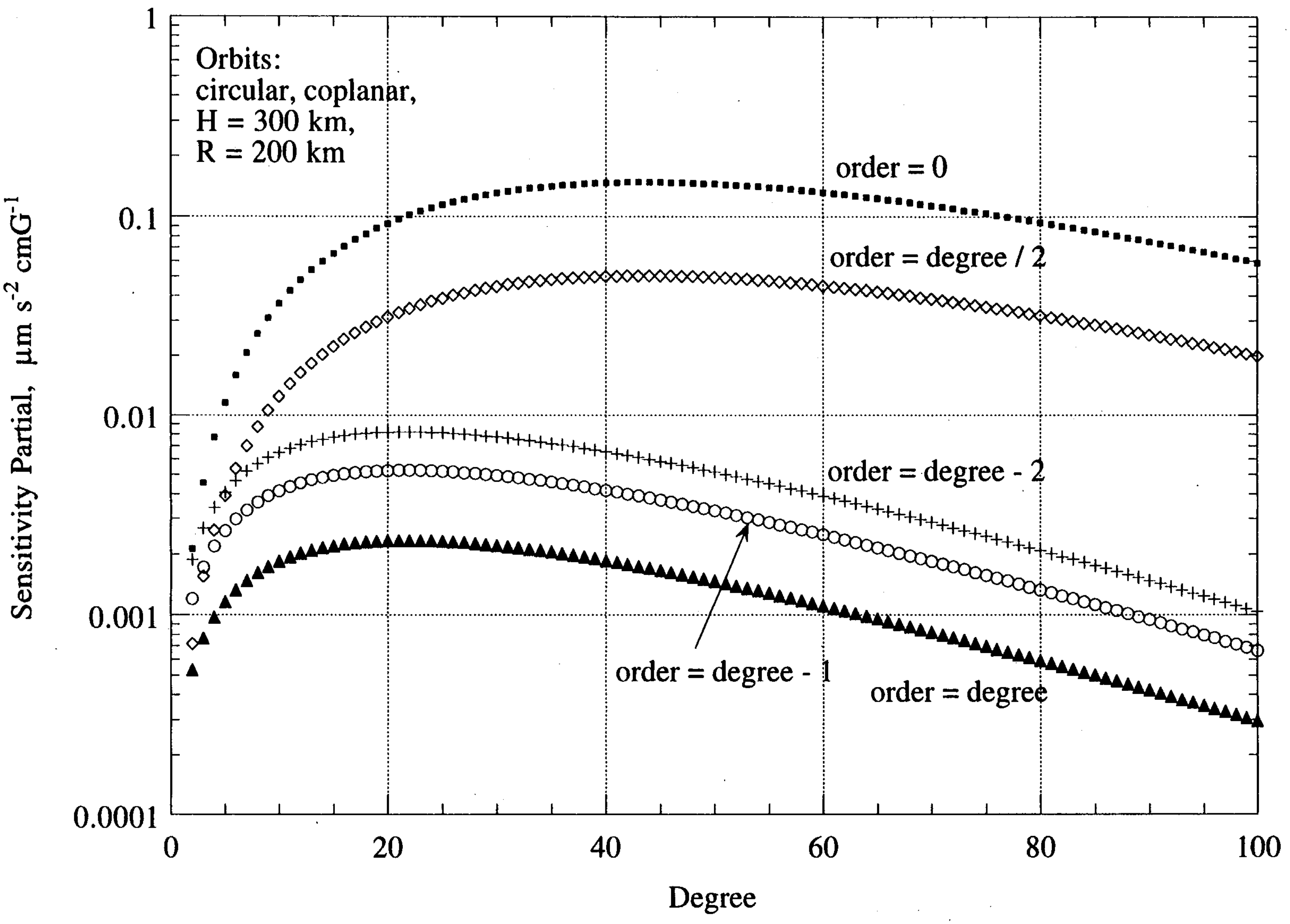

Figure 5-7 Approximate Sensitivity Partials for the Corrected Acceleration Observable

For the (n, 0) cases (zonals) in Figs. 5-8 through 5-11, tones (mainlobes) are located at n/rev, (n-2)/rev, (n-4)/rev, etc., but the dominant tone frequency is n/rev. This is a consequence of the fact that the dominant tone frequency in a (n, 0) spherical harmonic is its highest tone frequency at n/rev and the second derivative makes that tone even more dominant. Thus, the "effective frequency" for the partials for the (n, 0) harmonics is given by

$$f_{eff} \approx n/\mathrm{rev} \qquad (5.24)$$

In Figs. 5-9 and 5-10, for example, the amplitude of the n/rev tone is approximately 2.3 times larger the next largest tone at (n-2)/rev. Note in these plots that the width of the mainlobes, null to null, is $2/T_{DFT}$, where T_{DFT} is the somewhat arbitrarily selected integration time for the DFT.

The peaks between tones and higher than n/rev in these examples are the combined effect of sidelobes arising from the k/rev tones. These sidelobes decrease in amplitude relative to the tone mainlobes in proportion to DFT integration time (or fit filter integration time). In Fig. 5-9, for example, the sidelobe amplitude at 0.005 Hz is approximately 100 times smaller than the mainlobe at n/rev = 0.0037 Hz, as one would approximately calculate on the basis of frequency offset and integration time (i.e., 0.0013 Hz*4 *5400 s*π where integration time is 4 revs). If DFT (or fit-filter) integration time is increased to 13 days (208 revs), these sidelobes would decrease by an additional factor of 50 to 1/5000 relative to the tone mainlobe at 0.0037 Hz. Thus, even though the sidelobes in these plots are relatively large due to the relatively short integration time, in practice they are very small in actual fits.

Examples of the spectral content of the partials for (n, n) spherical harmonics (sectorials) are shown in Figs. 5-12 through 5-15. For comparison, each (n, n) harmonic is presented for two cases: fixed longitude and varying longitude (i.e., at 2π/sidereal-day). The fixed-longitude plots neglect the longitude dependence ($\cos(m\lambda)$ or $\sin(m\lambda)$) of the spherical harmonic and incorporate only the variation due to latitude change. As seen in the corresponding varying-longitude plots, inclusion of longitude dependence splits each latitude tone into two tones located approximately ±(1/16)*m/rev on either side of the given tone. Consideration of such (n, n) plots indicates that effective frequency for the acceleration observable can be crudely approximated by

$$f_{eff} \approx \sqrt{n}/\mathrm{rev} \qquad (5.25)$$

along with a warning that the frequency spread of the tones for each (n, n) harmonic is very broad, spanning a frequency range from nearly zero to twice the effective frequency.

These results show that the effective frequencies for the (n, 0) harmonics are much higher than for the (n, n) harmonics. In fits to the acceleration observable, for example, the (n, 0) harmonics will be particularly sensitive to noise at n/rev, while the (n, n) harmonics will be sensitive to noise closer to $\sqrt{n}$/rev. For the example of 5400 s per revolution, the effective frequencies are 18.5 mHz for (n, m) = (100, 0) and 1.85 mHz for (n, m) = (100, 100).

The preceding derivation for effective frequency implicitly assumes the spectrum of observable noise is nearly flat. When the noise spectrum is very strongly dependent on frequency, effective frequency can deviate from this flat-spectrum approximation, particularly for (n, n) harmonics. For the acceleration observable, accelerometer noise varies as 1/f in power. One can show for 1/f noise that the preceding approximations for effective frequency for acceleration partials are accurate for zonals and within a factor of 1.4 in terms of mapped noise for sectorials. System noise varies as f^2 (f^4 in power) for the acceleration observable. For such dependence, the effective-frequency approximation in Eq. (5.24) is quite accurate for zonals and, given the

15

effective "noise floor" shown in Fig. A-16, the effective frequency in Eq. (5.25) is a fair approximation for sectorials of lower degree.

Random errors and tone-like errors in observables map to geoid error in different ways in a fit that estimates gravity coefficients. Random noise mapped to geoid error for a given spherical harmonic can be approximately viewed as the sum of the noise power contributed by a number of bandpasses, each 1-μHz (for a 13-day fit) wide, centered on a tone found in the given partial, and scaled according to tone amplitude. Since bandwidth decreases in inverse proportion to fit interval, random noise errors in gravity coefficients decrease in proportion to $1/\sqrt{T_{fit}}$ as fit interval increases. On the other hand, in order for an error <u>tone</u> at a given frequency to be strongly mapped to geoid error, the error tone must closely "follow" a particular strong tone in the partial over the entire (13 day) fit interval with closely synchronous phase (e.g., $\Delta\phi < 0.1$ cyc). That is, when an error tone is sufficiently synchronous in frequency and phase with an important tone in a partial for a given gravity coefficient, the error caused by that tone in that coefficient does not average down as fit interval increases. On the other hand, if a tone is only slightly different in frequency, it can be highly suppressed. For example, a constant $\Delta f = 100$ μHz implies suppression by a factor of 300 (π 100 μHz/1 μHz) or more, in the case of a 1 μHz fit-filter bandwidth. Similar strong suppression occurs if the error tone wanders off in phase by a large amount with any other time behavior.

5.6 Crude Estimates of Geoid Error Based on Corrected Acceleration

The approximations for RMS partial and for effective tone frequency derived in the preceding two subsections can be used in Eq. (5.15) to estimate geoid error on the basis of the zero-correlation approximation. As indicated in Fig. 4-6, the highest-order harmonics (e.g., $m = n$ and $m = n-1$) have the smallest RMS partials for all degree values. Further, as suggested by Eqs. (5.24) and (5.25), these highest-order partials have the smallest effective frequencies. For accelerometer noise in Fig. B-1, the spectral noise density (SND) is largest at low frequencies. Thus, for accelerometer noise, the combination of small partial and low effective frequency will cause the two $m = n$ coefficients to dominate the errors for a given degree. (That is, as order decreases from $m = n$, partial magnitude increases, and effective frequency increases so that SND decreases, causing coefficient error to rapidly decrease relative to $m = n$.) Thus, to approximate per-degree geoid error for accelerometer noise, the RMS partials computed according to Eq. (5.22) for $m = n$ and effective frequency computed according to Eq. (5.25) are substituted in Eq. (5.15). ("Per-degree" geoid error for a given degree value is computed as the RSS of coefficient errors for all orders for that degree value.) This result is multiplied by $\sqrt{2}$ to account for the cos and sin cases and by another $\sqrt{2}$ to crudely account for other lower-order coefficients for each degree value. The resulting geoid error per degree, which is plotted in Fig. 5-16, decreases from approximately 30 μm for degree 2 to 5 μm for degree 20 through 50.

Geoid error caused by system noise is computed in a similar way, but in this case, the low-order coefficients make a much larger contribution to per-degree error. Note that the system-noise SND in Fig. B-1 increases rapidly as frequency increases, in proportion to f^2. Thus, as order decreases and effective frequency increases, the effective SND increases. This SND increase is sufficient to compensate for the increase in RMS partial as order decreases, making low-order coefficients contribute substantially to per-degree error. A crude analysis indicates that evaluating Eq. (5.15) for $m = n$ and then multiplying by $\sqrt{2}$ to account for sin and cos and by $1+3(n-1)/50$ to account for lower-order coefficients, approximately accounts for all orders for a given degree n, up to degree 50, after including the "noise floor" in Fig. A-16. The resulting geoid error per degree, which is plotted in Fig. 5-16, decreases from about 5 μm at degree 2 to about 2 μm at degree 10 and then increases to about 7 μm at degree 50.

Geoid error caused by USO noise can be computed through approximations similar to accelerometer noise since the increase in SND at higher frequencies is not sufficient to counteract the increase in partial magnitude. The resulting geoid error per degree, which is plotted in Fig. 5-16, decreases from about 2 μm at degree 2 to about 0.6 μm at degree 20 and then increases to about 1 μm at degree 50. Full simulations using range rate and estimating gravity coefficients and orbits have confirmed this crude estimate for USO error to within an order of magnitude [Srinivas Bettadpur, private communication, 1998].

As discussed in the next section, the crude geoid errors obtained above with the zero-correlation, closed-form approximation have been tested at the order-of-magnitude level for system noise and accelerometer noise with more complete multiparameter simulations based on hypothetical acceleration fits.

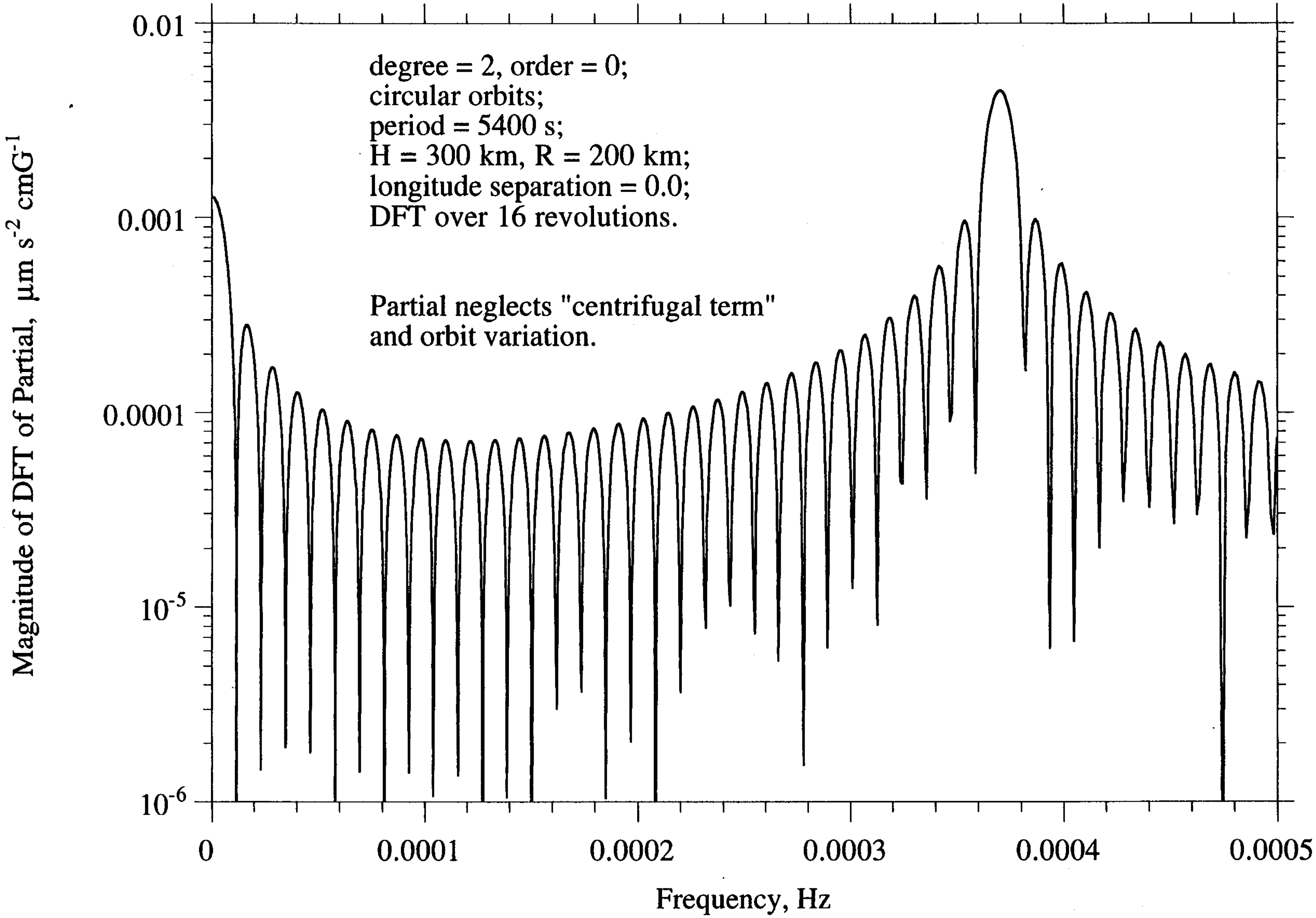

Figure 5-8 The Spectral Content of a Partial Derivative of Corrected Acceleration with respect to the (2, 0) Gravity Coefficient

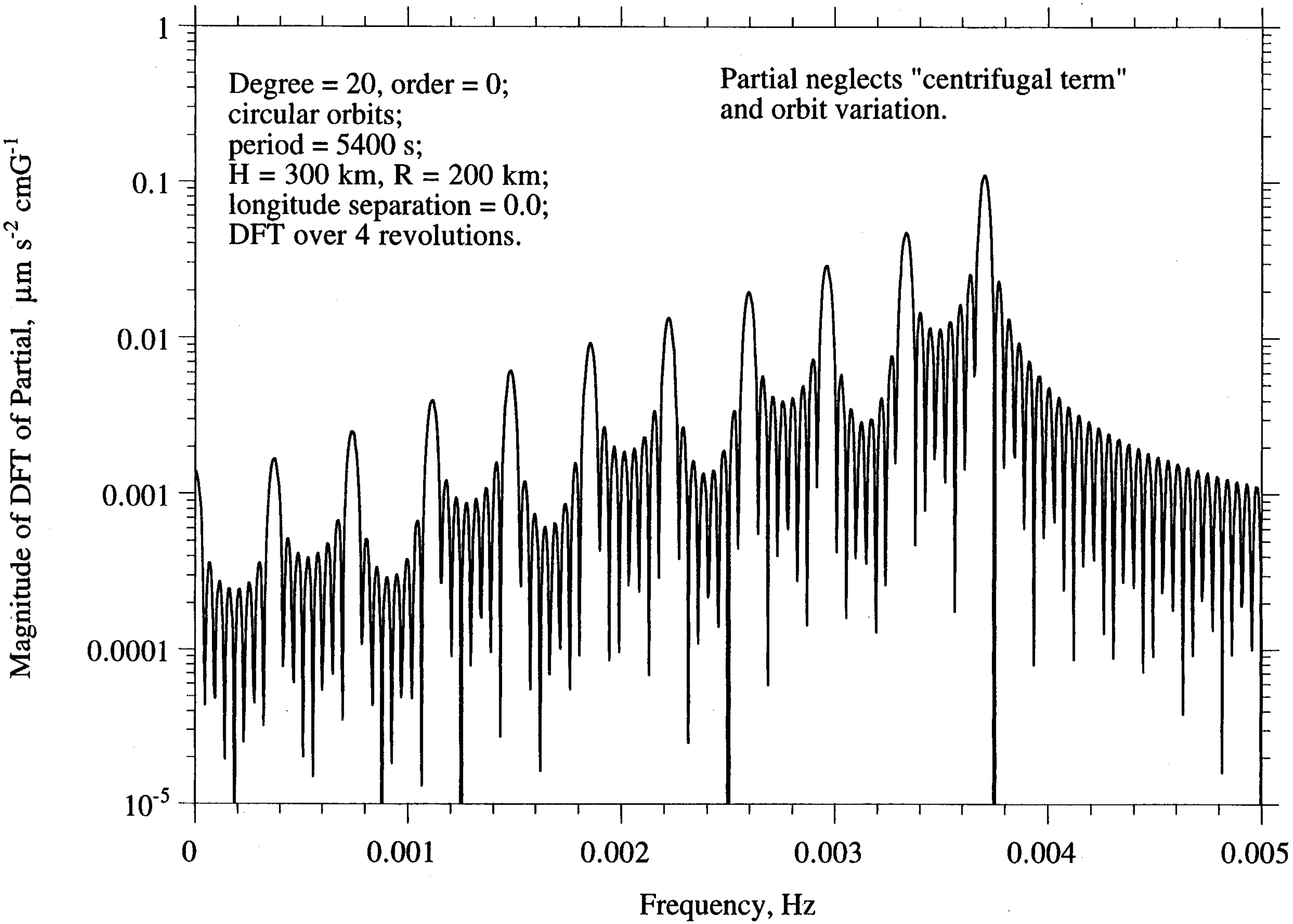

Figure 5-9 The Spectral Content of a Partial Derivative of Corrected Acceleration with respect to the (20, 0) Gravity Coefficient

10

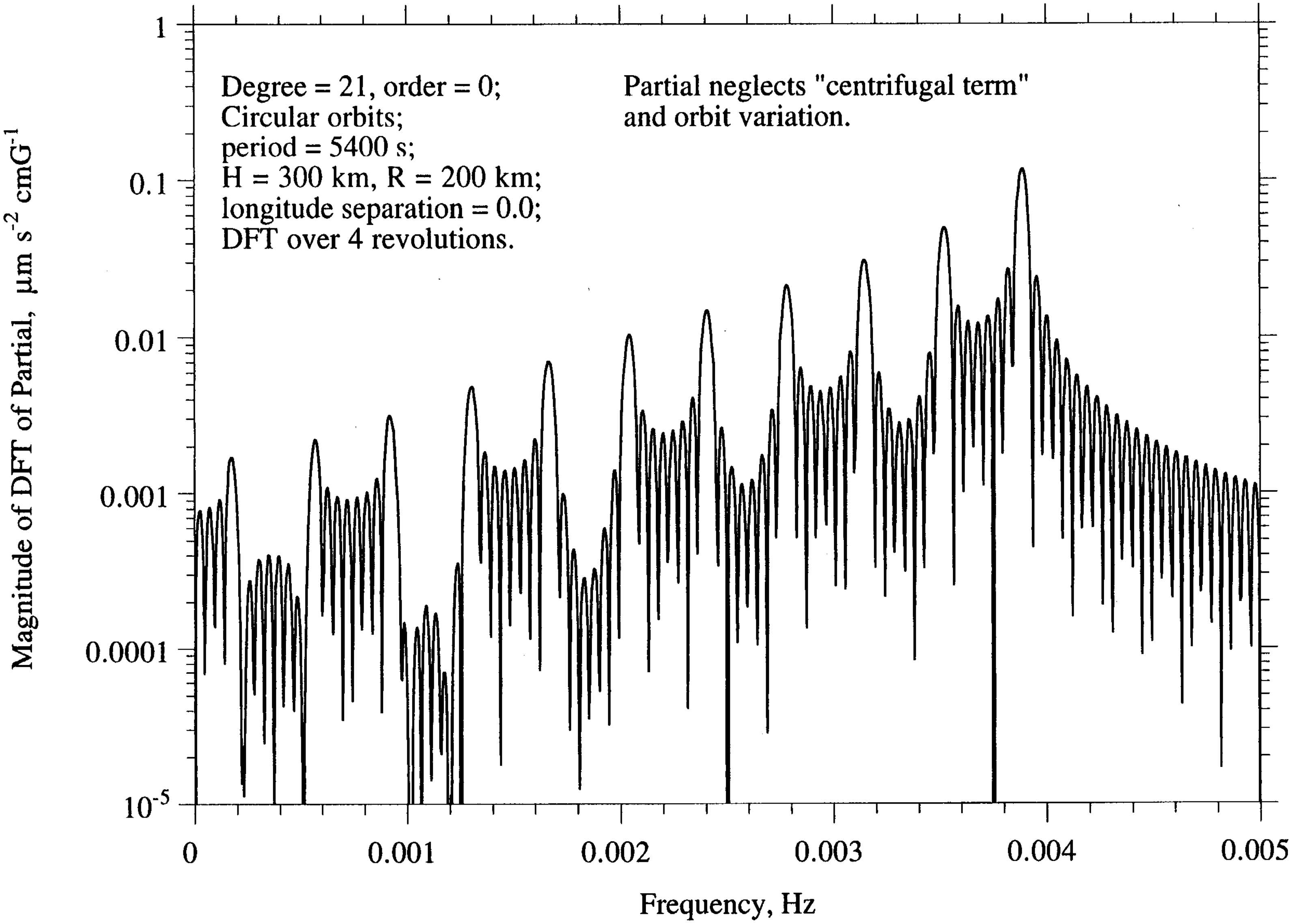

Figure 5-10 The Spectral Content of a Partial Derivative of Corrected Acceleration with respect to the (21, 0) Gravity Coefficient

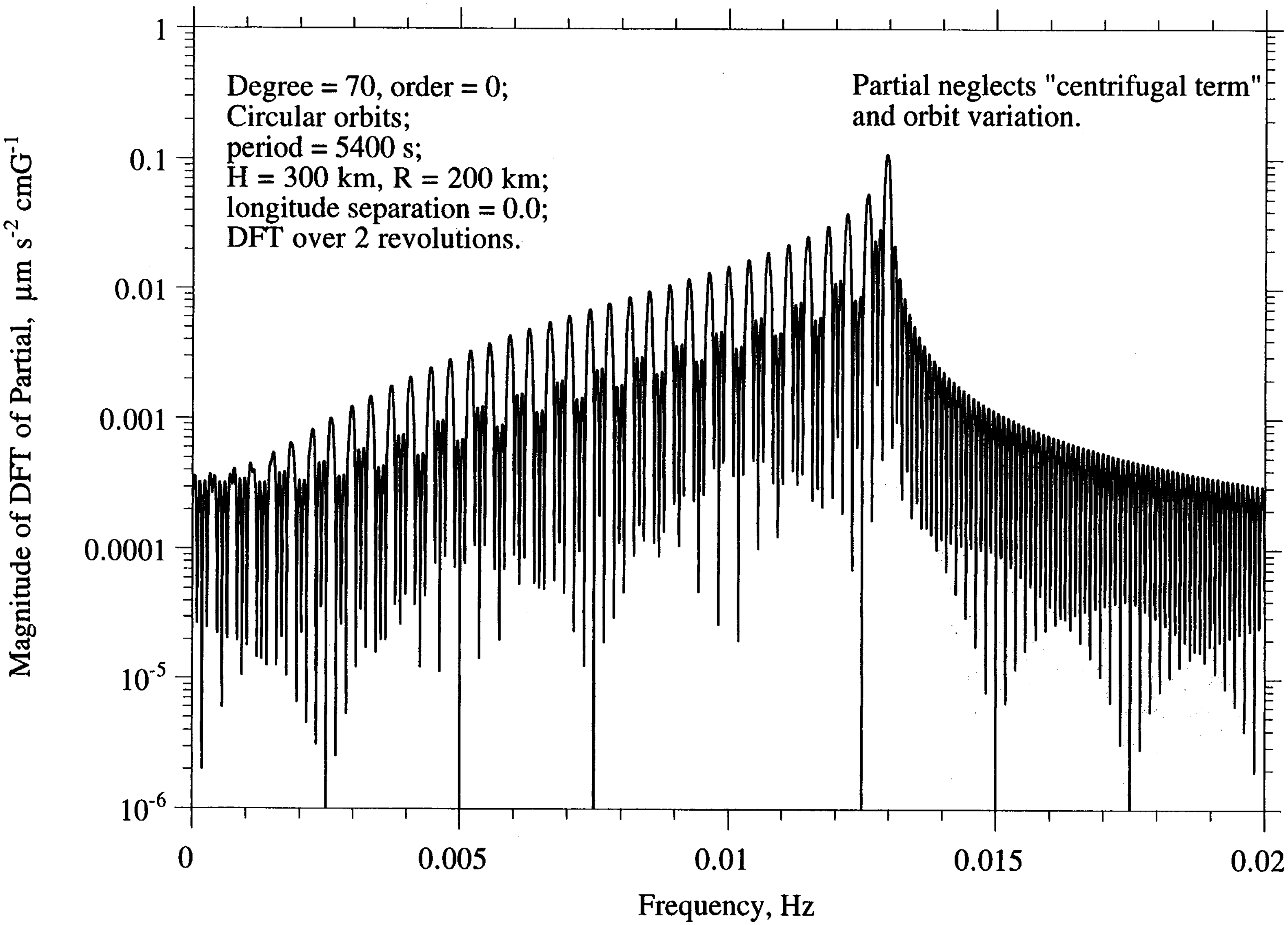

Figure 5-11 The Spectral Content of a Partial Derivative of Corrected Acceleration with respect to the (70, 0) Gravity Coefficient

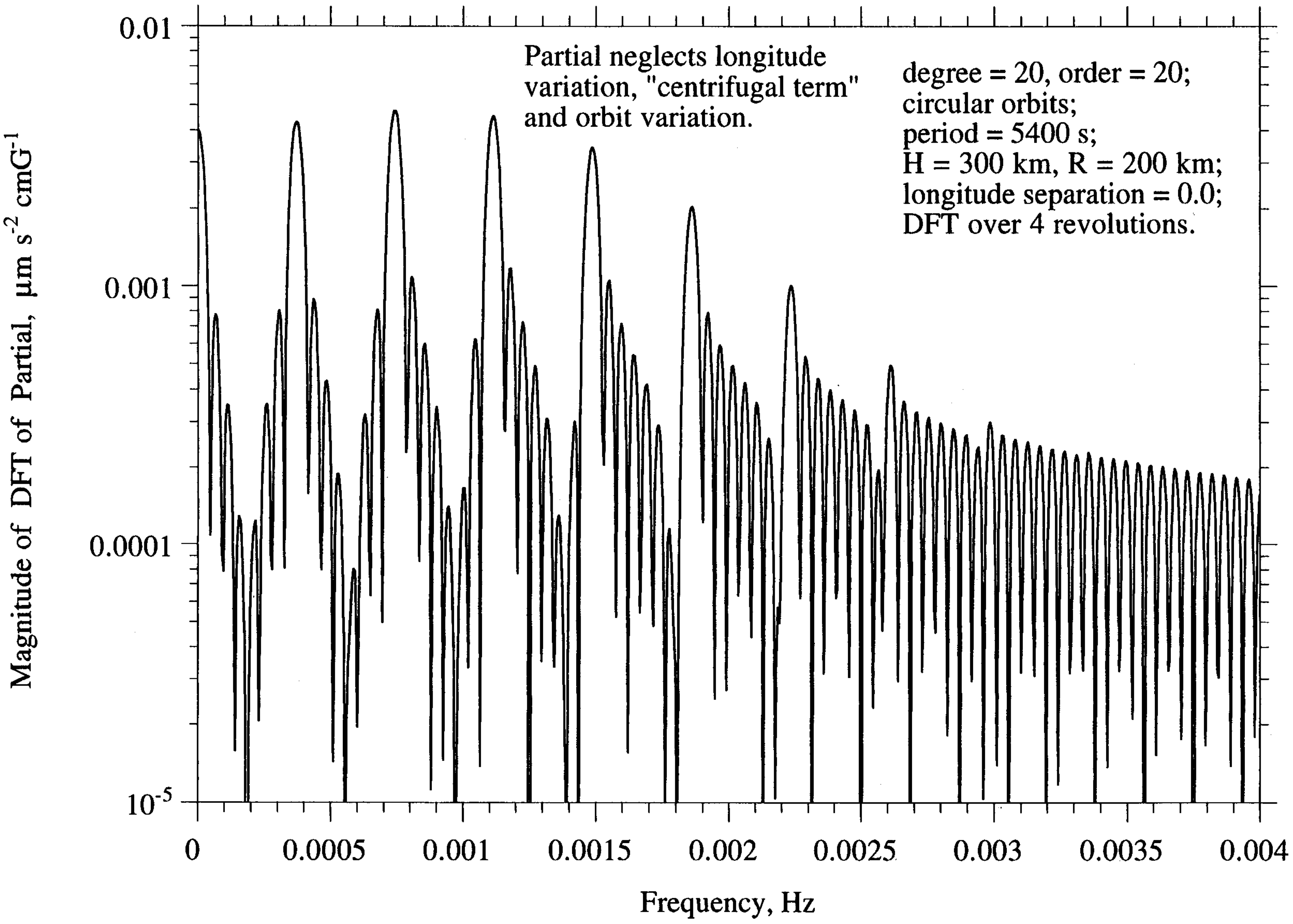

Figure 5-12 The Spectral Content of a Partial Derivative of Corrected Acceleration with respect to the (20, 20) Coefficient, Case A

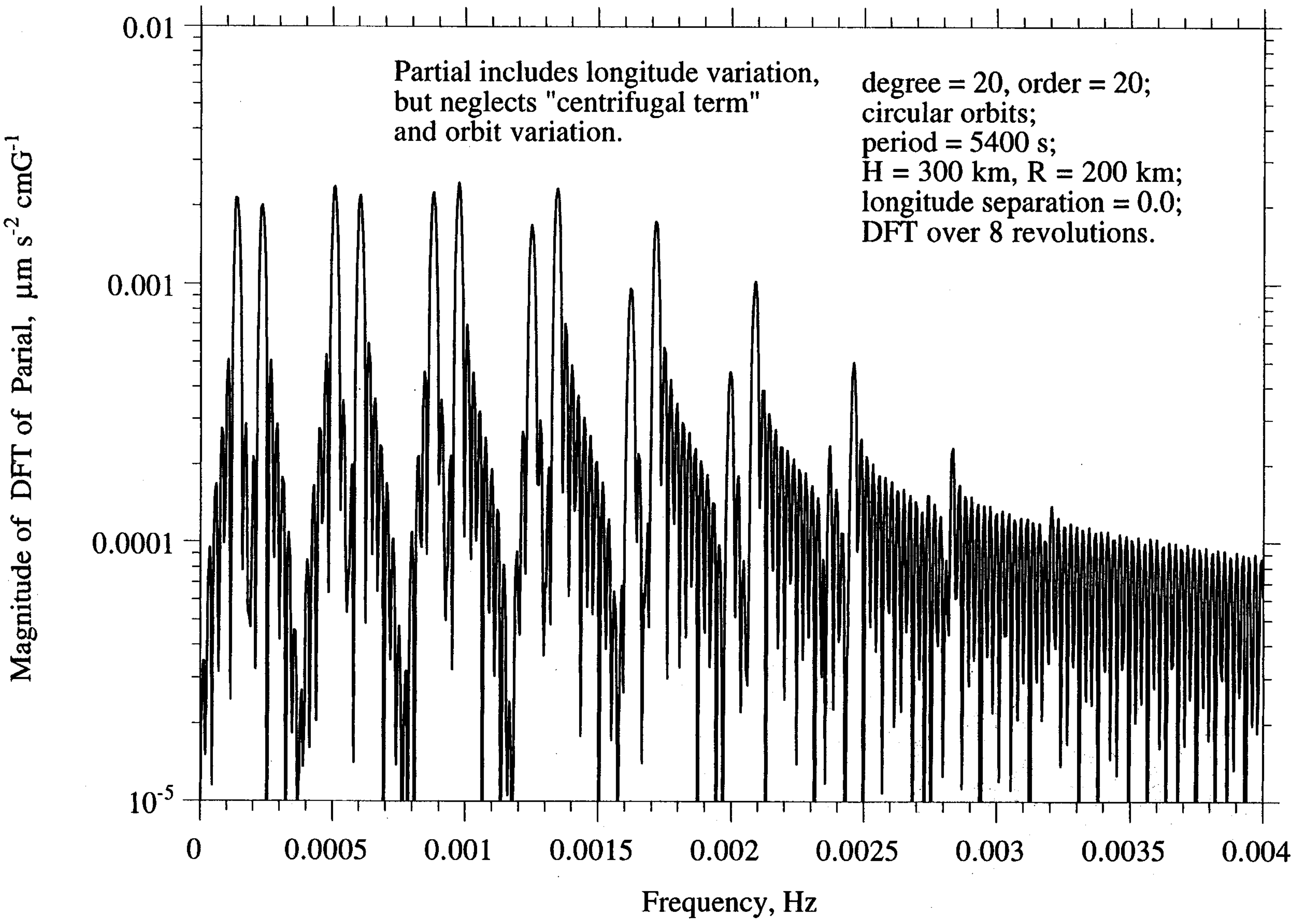

Figure 5-13 The Spectral Content of a Partial Derivative of Corrected Acceleration with respect to the (20, 20) Coefficient, Case B

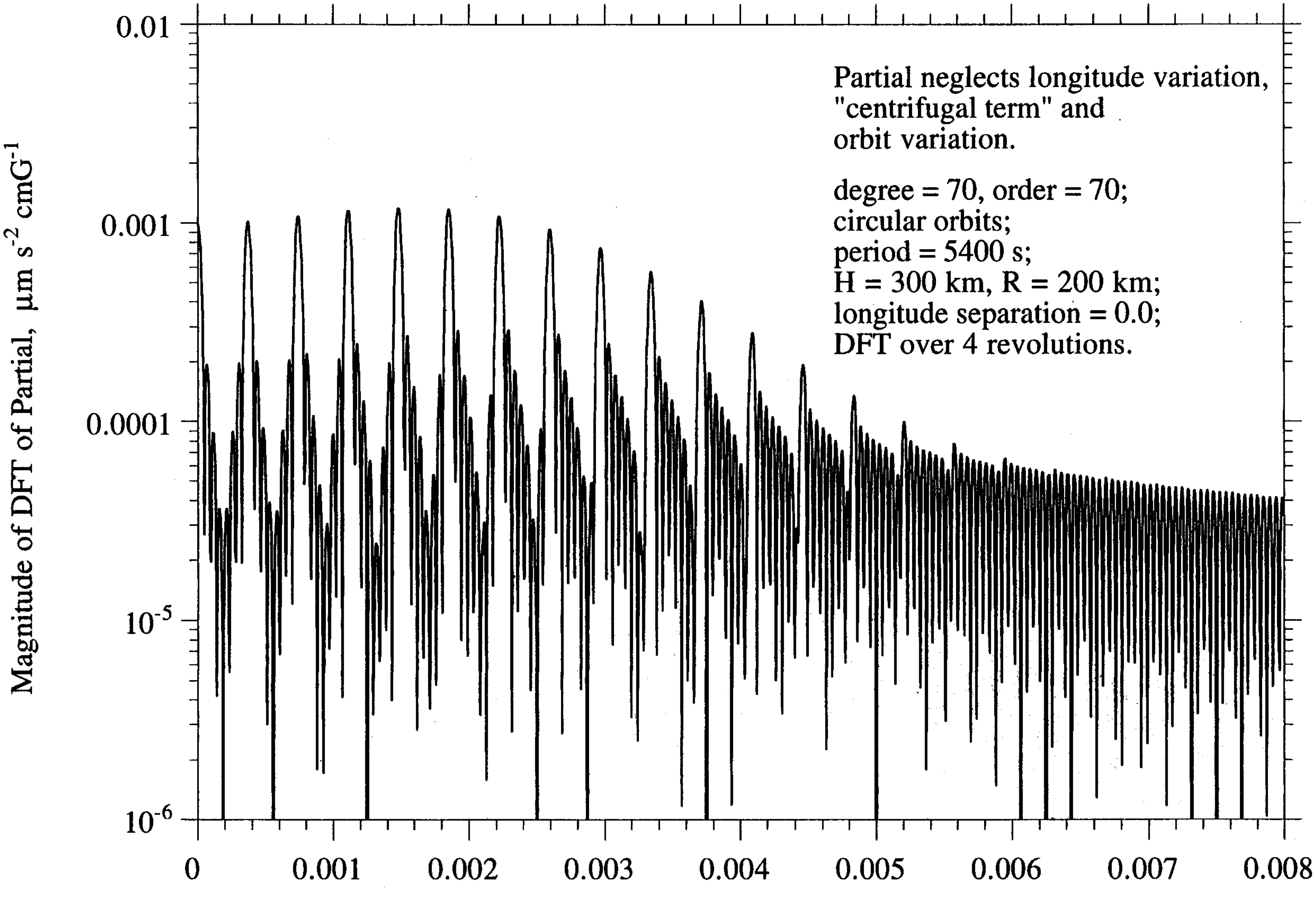

Figure 5-14 The Spectral Content of a Partial Derivative of Corrected Acceleration with respect to the (70, 70) Coefficient, Case A

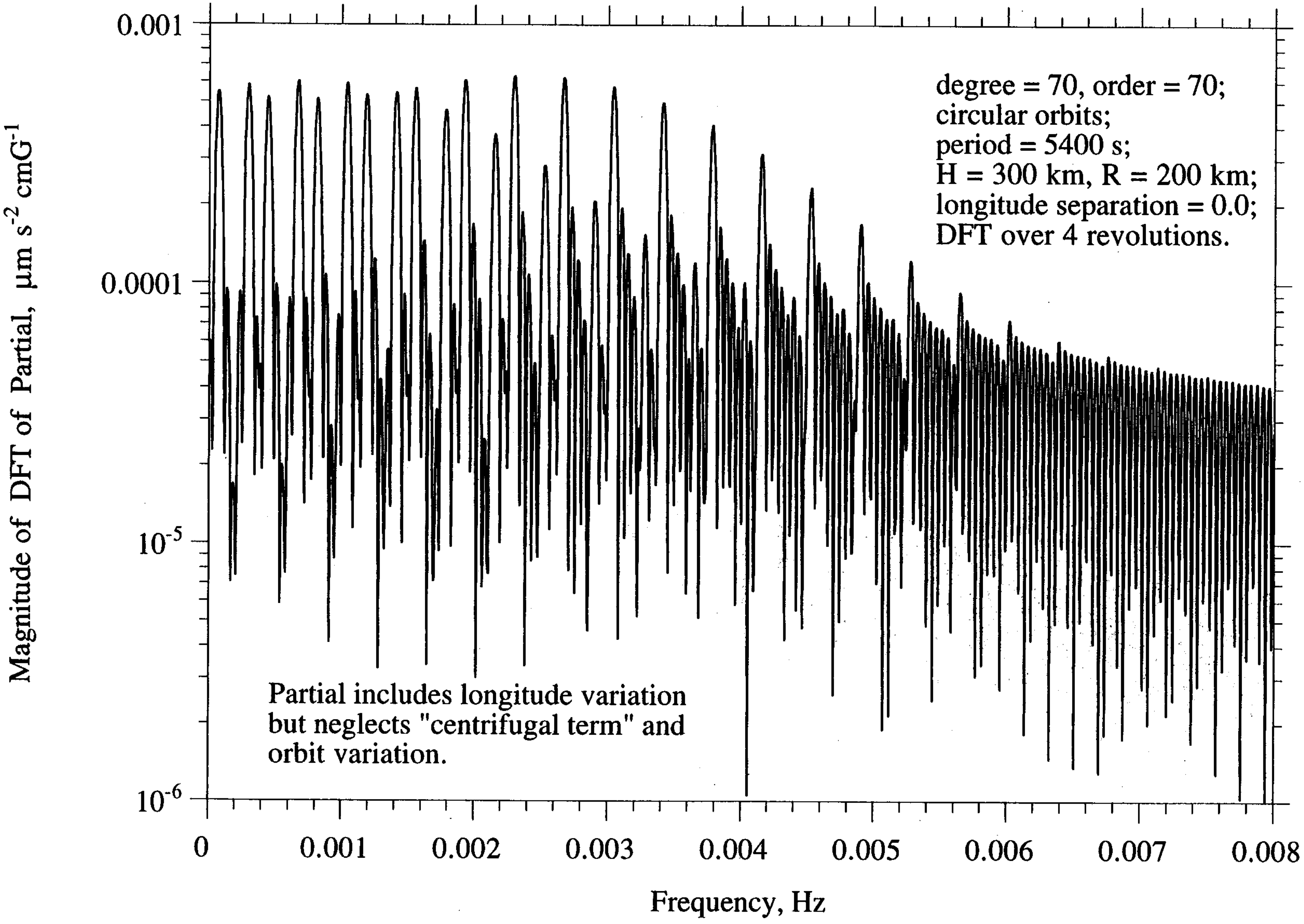

Figure 5-15 The Spectral Content of a Partial Derivative of Corrected Acceleration with respect to the (70, 70) Coeffficient, Case B

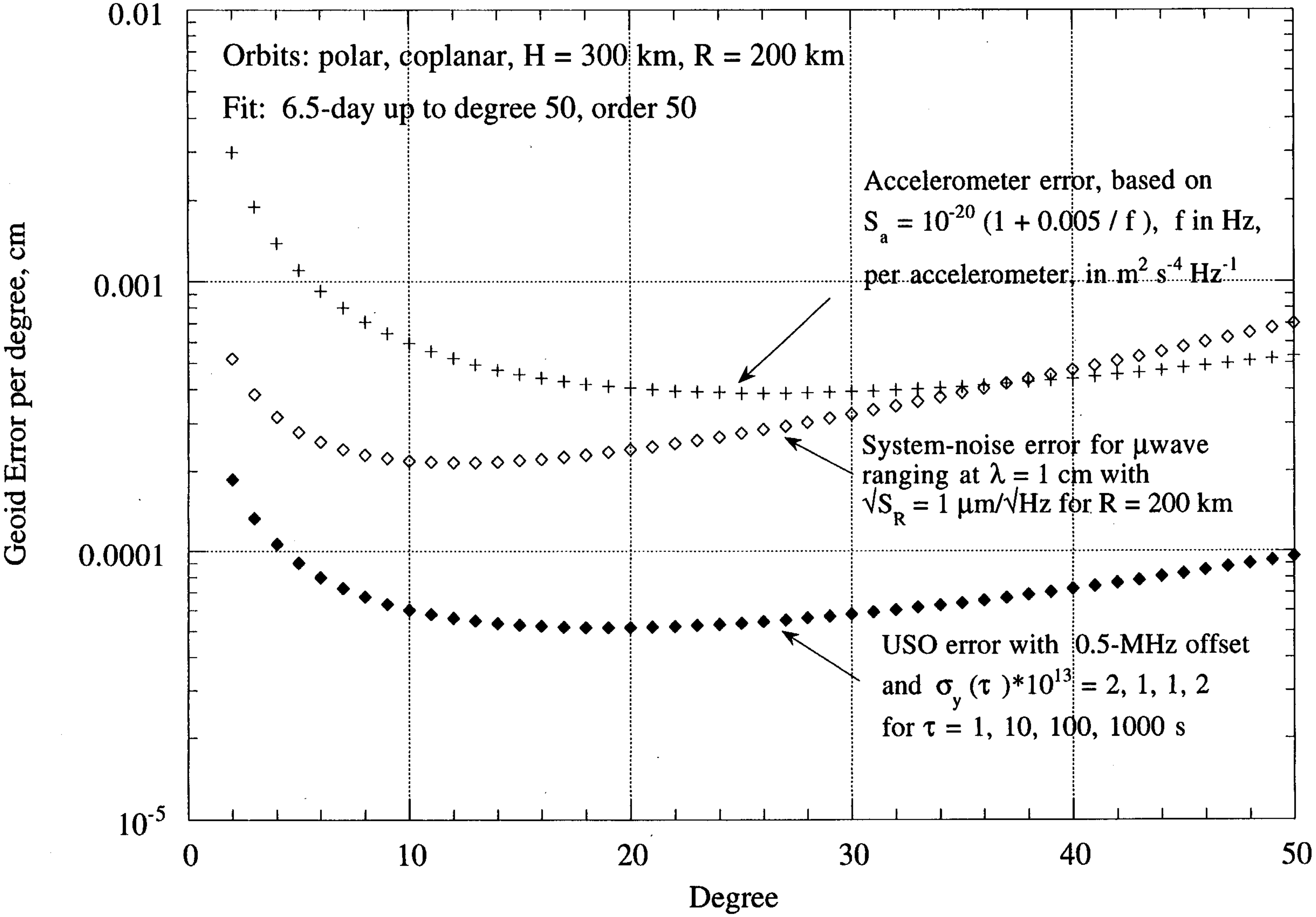

Figure 5-16 Gravity Coefficient Errors from a Hypothetical Fit to the Corrected Acceleration Observable, Computed with the Approximate Closed-Form Expression

SECTION 6

ERROR SIMULATIONS WITH FITS TO CORRECTED ACCELERATION

6.1 The Adopted Approach for Simulations

In order to obtain more accurate results, this section carries out multiparameter solutions that do not contain the zero-correlation approximations of Section 5. Two types of random errors, system noise and accelerometer noise, have been simulated and mapped to gravity-coefficient errors for the case of a hypothetical fit to the corrected acceleration observable. For these simulations, the simultaneously "estimated" parameters include only the gravity coefficients and instrumental parameters. Since the observables are linear with respect to all estimated parameters under the assumptions that the orbits are known and the centrifugal term is "corrected," partial derivatives of the acceleration observable with respect to gravity coefficients can be computed as described in Section 4.

A number of fitting methods can be used to extract gravity coefficients. The least-squares method adopted in this report is a sequential U-D filter that updates the solution one (10-s) observation at a time, in sequence of observation, until the end of the fit interval (e.g., 6.5 days with 56,000 observations) is reached. To provide stability, the fit is initialized by assigning a large initial error (e.g., 10^5 to 10^7 cmG for gravity coefficients) to each of the parameters to be estimated. As the fit is updated one observation at a time, each input observable is assigned the same nominal error weight, one that approximates the total error on each observable (e.g., in the range 10^{-4} to 10^{-2} $\mu m/s^2$). In principle, final values for the estimated parameters depend on the ratio of the initial parameter error and the assigned observable error. However, the range of values allowed for the initial parameter error and the observable weighting error is wide since estimated parameters settle at the end of the fit interval to essentially the same values for a wide range of ratio values (i.e., over orders of magnitude). The adopted fitting method does not account in the fit for possible correlations between noise on input observables (as simulated for accelerometer noise) and, in terms of final estimated parameters, is essentially the same as a conventional $A^T A$ least-squares fit (i.e., with a weighting matrix equal to a unit matrix that is "optimal" for uniform-variance observable errors, uncorrelated observation to observation).

In the simulations, the bias term a_c and the "signal" part of each intersatellite acceleration observable, $(\mathbf{a}_2 - \mathbf{a}_1) \cdot \hat{\mathbf{R}}$, are set to zero (i.e., all gravity coefficients and the point mass term are all effectively set to zero) so that the input "observables" consist only of the particular error being considered. Including the signal part would add no new information to the error analysis since the signal would be removed by the fit in exactly the same way it was incorporated, leaving the same output error. (This is equivalent to assuming the initial model is perfect so that observed-minus-computed (OMC) is zero except for the error term.) Since the observations are linear in the estimated parameters in the assumed approach, as discussed in Section 4, this approach does not compromise the analysis with respect to convergence considerations. In this error analysis, only one computation of partials and one "fit" are necessary.

In order to separately assess the contribution of each error source, each simulated error source is separately passed through the sequential filter. Both types of errors—random errors, generated by shaping the frequency spectrum of the output of a white-noise random number generator, and systematic errors, tailored to match respective error models—can be input to the sequential filter as a numerical sequence of "observables" in order to determine their respective effect on estimated parameters. Only results for random errors are presented in this report. In the case of random errors, a given fit provides only the response of the fit to the particular numerical sequence generated for that fit (i.e., one member of the ensemble of possible noise sequences), and therefore represents only one response out of many possible responses. In principle, many separate fits, each with a different input noise sequence (seed) are required to estimate the distribution and 1-σ error for each estimated parameter. However, for GRACE, so many coefficients are estimated (e.g., approximately 2600 for degree 50) that coefficient scatter can be fairly accurately estimated for most gravity coefficients, after using only one seed, by looking at the scatter of adjacent coefficients (i.e., with nearly the same degree and order). An exception to this approach is the low-degree harmonics for which the order count is small (e.g., only five coefficients for degree = 2). These low-degree terms can require more than one seed if high-accuracy error estimates are needed.

6.2 Satellite Orbits

In the calculation of the partials in these simulations, orbit positions for the two satellites are each computed from the simple closed-form expressions that generate an exactly circular, polar (90° inclination) orbit, with selectable radius and with selectable separation between satellites, in both latitude and longitude. The two orbits are assigned the same radius to enforce orbit synchronism and continuous visibility between satellites. More sophisticated orbits incorporating gravity harmonics are not necessary in this analysis, since exactly circular orbits can provide the most important features required for error analysis. For example, circular orbits are adequate with respect to sampling of the lobes of the partials, in both latitude and longitude, including longitude commensurability effects. More realistic orbits can shift the placement of sample points within lobes and slightly change orbit radius, but should not significantly change overall results, given adequate lobe sampling density (e.g., at least two well-spaced samples per lobe for the best orbits). For example, the largest gravity term, J2, causes 2/rev variations per satellite on the order of a km relative to a circular orbit. For the J100 harmonic (degree 100, order 0), the wavelength of the highest-frequency harmonic is approximately 400 km along track. Thus, J2 variations in orbit position are on the order of 0.003 harmonic cycle in this example and should not be important in sampling considerations. The orbit variations caused by other harmonics are much smaller than J2 and should cause even smaller alterations in the sampling distribution. Similar comments apply to sampling in longitude, eccentricity and orbit mismatch. Thus, the detailed variations found in actual orbits can be neglected in the error analysis presented here, since only moderately realistic sampling in latitude and longitude is required.

6.3 Simulation Results for Two Random Error Sources

Simulations using the sequential filter have been carried out for two error sources: system noise (thermal noise) and accelerometers. Because of computing-time limits, fits were carried out only up to (degree, order) = (50, 50). Given uniform sampling in longitude, a fit interval of 6 to 7 days can provide up to approximately four uniformly-spaced samples for each cycle of harmonic phase in longitude for the highest order terms, i.e., up to order 50. One constant instrumental parameter is estimated for the entire fit interval. (Expanding the instrumental model to a piecewise constant function does not affect results significantly for these simulations, as discussed in Appendix C.) In a 50x50 fit, 2600 gravity coefficients are estimated for a total of 2601 parameters.

Computation time for these fits is approximately 30 hours on a work station. Three orbit configurations have been simulated based on the following cases for satellite separations in (latitude, longitude) in km: (200, 0), (200, 40), (200, 200). All cases are assigned the same satellite altitude of 300 km with a period of 5430 s.

6.3.1 System Noise

Input "acceleration observables" for system noise are obtained by using a random number generator to generate "white" (i.e., uncorrelated from point to point) range noise at 10 S/s and then passing this noise through a double-differentiating low-pass digital filter (see Appendix A) that is designed to extract acceleration from actual intersatellite range observables. As discussed in Appendix A, the purposes of this filter are to decimate the data rate from 10 S/s to a selected output rate (0.1 S/s in this simulation), apply a second derivative, and low-pass filter with a selected bandwidth (0.05-Hz in this simulation), all in one step. The effect of this filter from the system-noise perspective is to convert white noise to "blue" noise with a f^2 frequency dependence (in "voltage"). The noise sequences generated with this approach should model actual system noise very accurately.

Based on the K-band SNR adopted here for an intersatellite range of 200 km (with $C/N_0 = 69$ dB-Hz), the spectral noise density is approximately 1 μm/$\sqrt{\text{Hz}}$ for dual-band-calibrated range (see Appendix B). After differentiation with the digital filter, the spectral noise density at frequency f for the resulting acceleration noise is given by $(1\ \mu\text{m} / \sqrt{\text{Hz}}) * (2\pi f)^2$ for frequencies between approximately f = 2 mHz and 20 mHz. Below f = 2 mHz, as discussed in Appendix A and shown in Fig. A-17, the fitting process effectively causes a flat noise floor at a level of approximately $10^{-4}\ \mu\text{m s}^{-2}\ \text{Hz}^{-1/2}$, given a fit interval of 13 days. Thus, given the fit bandwidth of approximately 1 μHz associated with a 13-day fit interval, the effective system-noise error for the acceleration observable within the fit bandwidth covers the range from $10^{-7}\ \mu\text{m s}^{-2}$ for signal tones with f < 1 mHz, up to $1.3\times10^{-5}\ \mu\text{m s}^{-2}$ for signal tones at f = 18 mHz (e.g., for n = 100, m = 0).

Results for a particular random-number-generator seed are presented in Figure 6-1 for the three orbit-separation cases cited above, with orbital altitudes of 300 km and orbital periods of 5430 s. This period is expected to be one of the best with regard to longitude sampling since it is close to the "uniform-sampling" period of 5433 s corresponding to a "7-day-repeat" commensurability ratio of 111/7 between orbital period and sidereal day. Also plotted in Figure 6-1 are errors predicted by the approximate error propagation analysis of Section 5 that assumes no correlations between estimated parameters. Plotted errors are "per degree," computed as the root-sum-square of the errors from all orders associated with a given degree.

As Figure 6-1 shows for the (200 km, 0 km) orbits, the geoid error per degree falls in the range 0.0002 to 0.0004 cmG for degree 2 to degree 10 and increases to approximately 0.0007 cmG for degree 50. Results from approximate error propagation of Section 5 for the (200 km, 0 km) case are in reasonable agreement with the simulation results except for degree 2. The degree-2 discrepancy is explained mainly by underestimation of the partial magnitude for degree 2 in the approximate error propagation analysis. For the (200 km, 40 km) case, geoid errors are only slightly better than the (200, 0) case (by approximately 25%). For the (200 km, 200 km) case, the geoid error per degree ranges from approximately 0.00005 cmG for low degree values to approximately 0.0004 cmG for degree 50. Comparison of the results for the (200 km, 0 km) orbits and (200 km, 200 km) orbits shows that increasing the longitude separation from 0 to 200 km improves the geoid error due to system noise by approximately a factor of 4 near degree 10 and a factor of two near degree 50, for these hypothetical acceleration fits. This improvement is a consequence of the stronger partials for harmonics of high order (i.e., m near n), as discussed in Section 4.

79

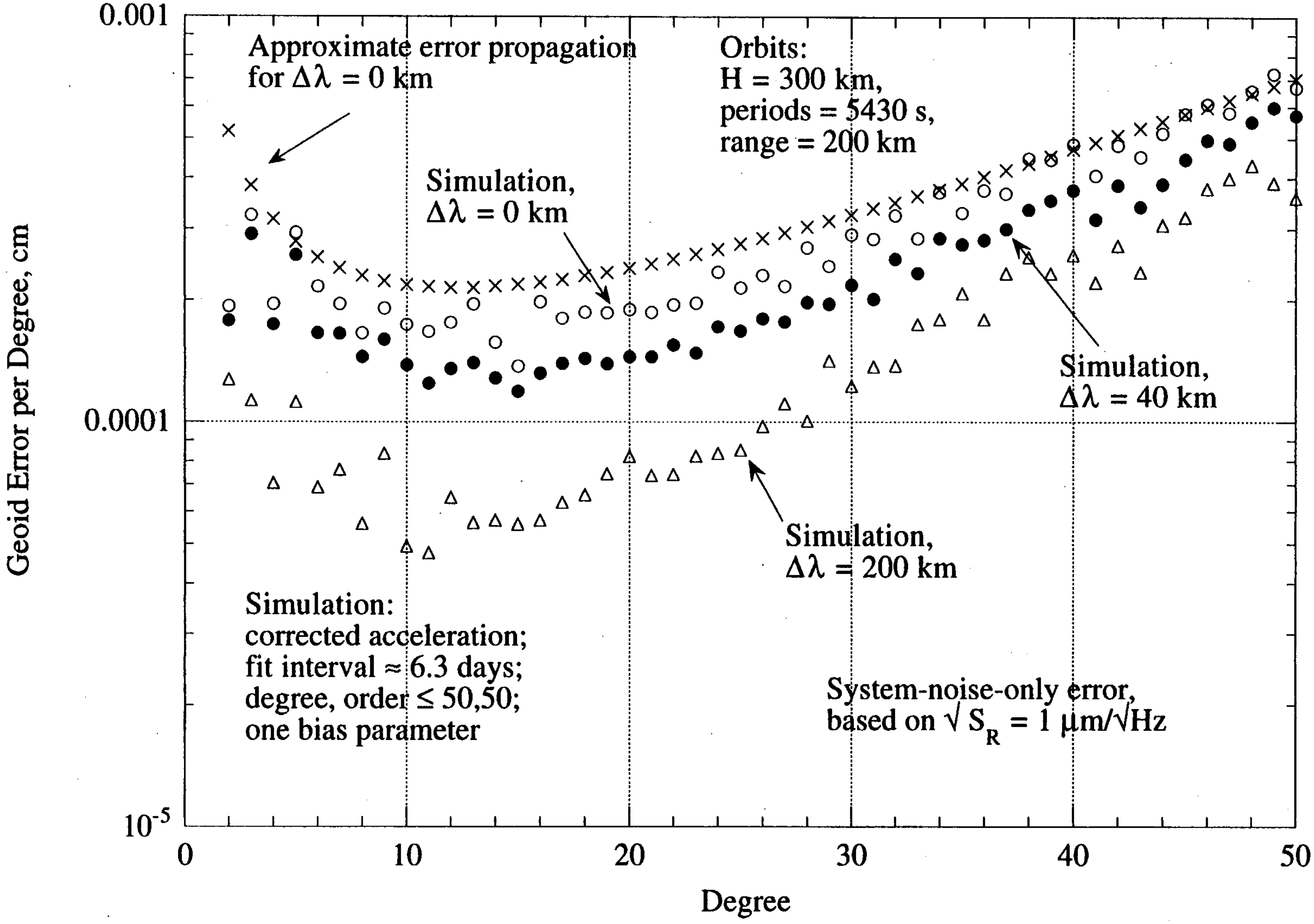

Figure 6-1 Example Simulations Of Geoid Error Caused by System Noise in Corrected Acceleration Observable

6.3.2 Accelerometer Random Noise

As discussed in Appendix A, the random error generated by the accelerometers can be simulated by passing white noise sampled at 10 S/s through a specially designed time-domain digital filter. This filter processes the white noise to apply a low-pass filter, decimate the sample rate from 10 S/s to 0.1 S/s, and shape the spectrum within the passband. (To shape the spectrum, one follows the approach described in Appendix A for the range filter, but the amplitudes in the passband are set to appropriate values in the frequency domain before transforming to the time domain. For purposes of noise generation, it is better to use an approximate Fourier transform rather than a strict discrete Fourier transform, since more points can be placed in the passband.) The simulated noise spectrum generated for the accelerometers, which is shown in Figure 6-2, matches the anticipated noise spectrum (from accelerometer specifications) fairly accurately above 100 μHz. The specification noise spectrum (power) along each of the two sensitive axes (line-of-sight and radial) is proportional to $(1 + 0.005/f)$ where frequency f is in Hz. Thus, the shape of the random noise spectrum is dramatically different for system noise and accelerometer. The power spectrum of accelerometer random noise is biased toward low frequencies in proportion to $1/f$, while system noise is biased toward high frequencies in proportion to f^4. This difference in power spectra results in substantial differences in associated geoid errors.

Each noise sequence generated by the noise-shaping digital filter can be separately processed by the sequential filter in a manner similar to that described for system noise above. Geoid errors obtained for a particular seed are plotted in Figure 6-3 for the orbits described above. For the (200 km, 0 km) orbits, the geoid error per degree falls in the range 0.001 to 0.0005 cmG for degree 2 to degree 10 and is approximately 0.0005 cmG for degree 10 to 50. Results from approximate error propagation for the (200, 0) case are in reasonable agreement with the simulation results, given the approximations of the error-propagation analysis. The geoid error for degree 2 is again too high by about a factor of two due to the inaccuracy of the approximate equation for the closed-form partials.

For the (200 km, 40 km) case, geoid errors are somewhat better than the (200, 0) case (by approximately a factor of 1.25 to 2). For the (200 km, 200 km) case, the geoid error ranges from approximately 0.001 cmG for the lowest degree values down to approximately 0.00006 cmG for higher degree values. Comparison of the results for the (200 km, 0 km) orbits and (200 km, 200 km) orbits shows that increasing the longitude separation from 0 to 200 km improves the geoid error per degree by approximately a factor of 10 for degree values between 20 and 50, for these hypothetical acceleration fits. The improvement is a consequence of the stronger partials for harmonics of high order (i.e., m near n), as discussed in Section 4. This hypothetical simulation for 200-km longitude separation does not correctly represent the effect of accelerometer noise when the "out-of-plane" axis of the accelerometers is much noisier than the other axes. A noisier "out-of-plane" axis can substantially degrade performance in the (200 km, 200 km) case.

Since geoid errors are presented as per-degree errors, Figure 6-3 does not show the dramatic improvement experienced by the high-order coefficients when longitude separation is increased from 0 to 200 km. For the fits described above, this improvement is illustrated in Figures 6-4 and 6-5, in which coefficient errors are presented as a function of order for degree 50. Note that the errors for the low values for order do not change substantially while the errors for the highest orders improve by approximately a factor of 15 when longitude separation is increased from 0 to 200 km.

One potential shortcoming of these accelerometer simulations is that the simulated pseudorandom noise does not continue to diverge as strongly as $1/f$ at the very lowest frequencies (i.e., $f < 0.1$ mHz). As suggested by the analysis in Appendix C, however, better noise simulations containing such divergence should not change results when instrumental noise is properly modeled.

These hypothetical-fit results further emphasize the advantage provided by larger longitude separation. Because of disadvantages such as the increased attitude/pointing demands and increased susceptibility to accelerometer-weak-axis noise, however, a 200-km separation does not appear to be feasible, at least at this time. If a longitude separation of 0 to 40 km is implemented, differences in partials will lead to a large imbalance between errors for low-order and high-order gravity coefficients, outlined in Section 4.

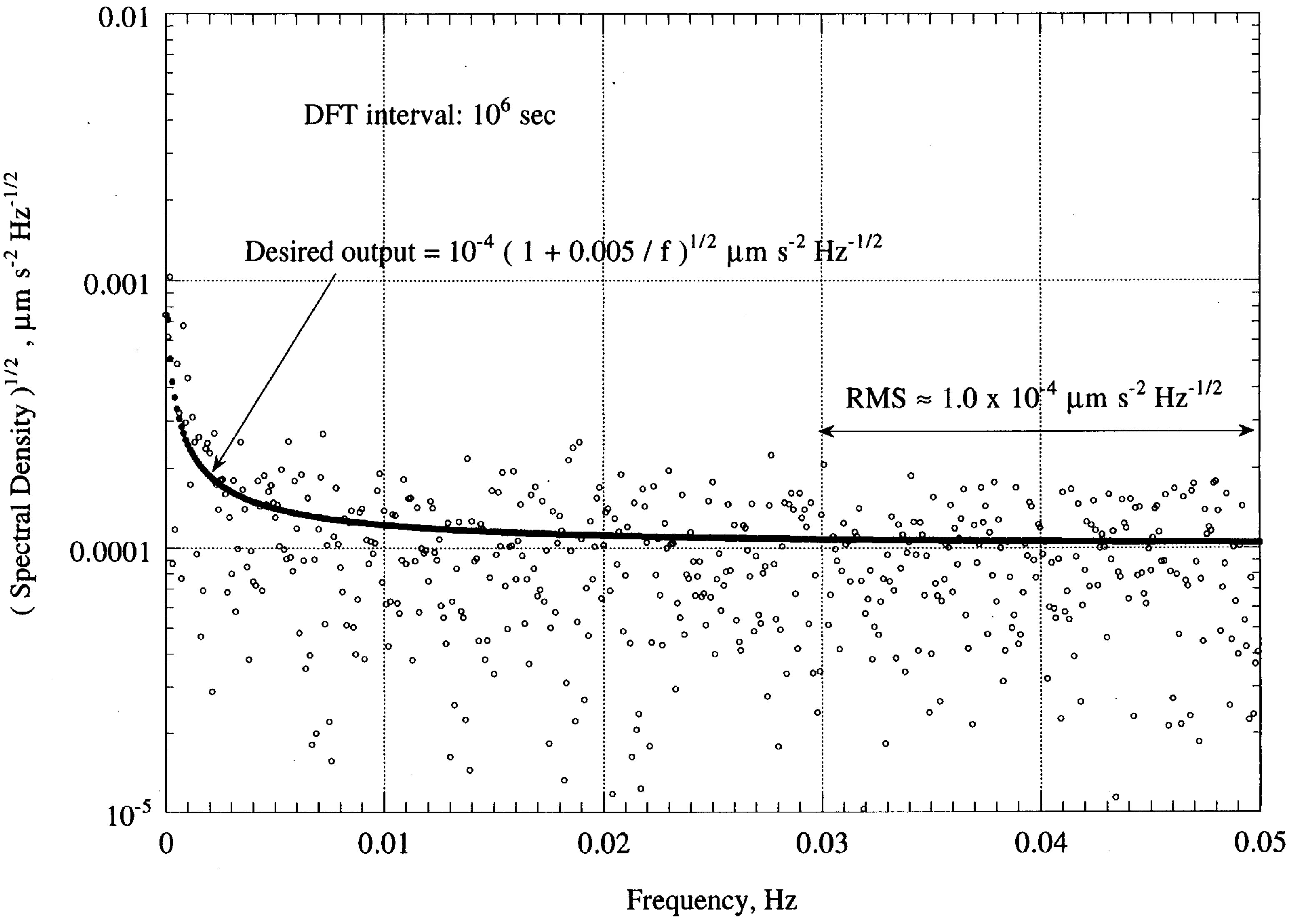

Figure 6-2 DFT of a Noise Sequence that Simulates Accelerometer Noise

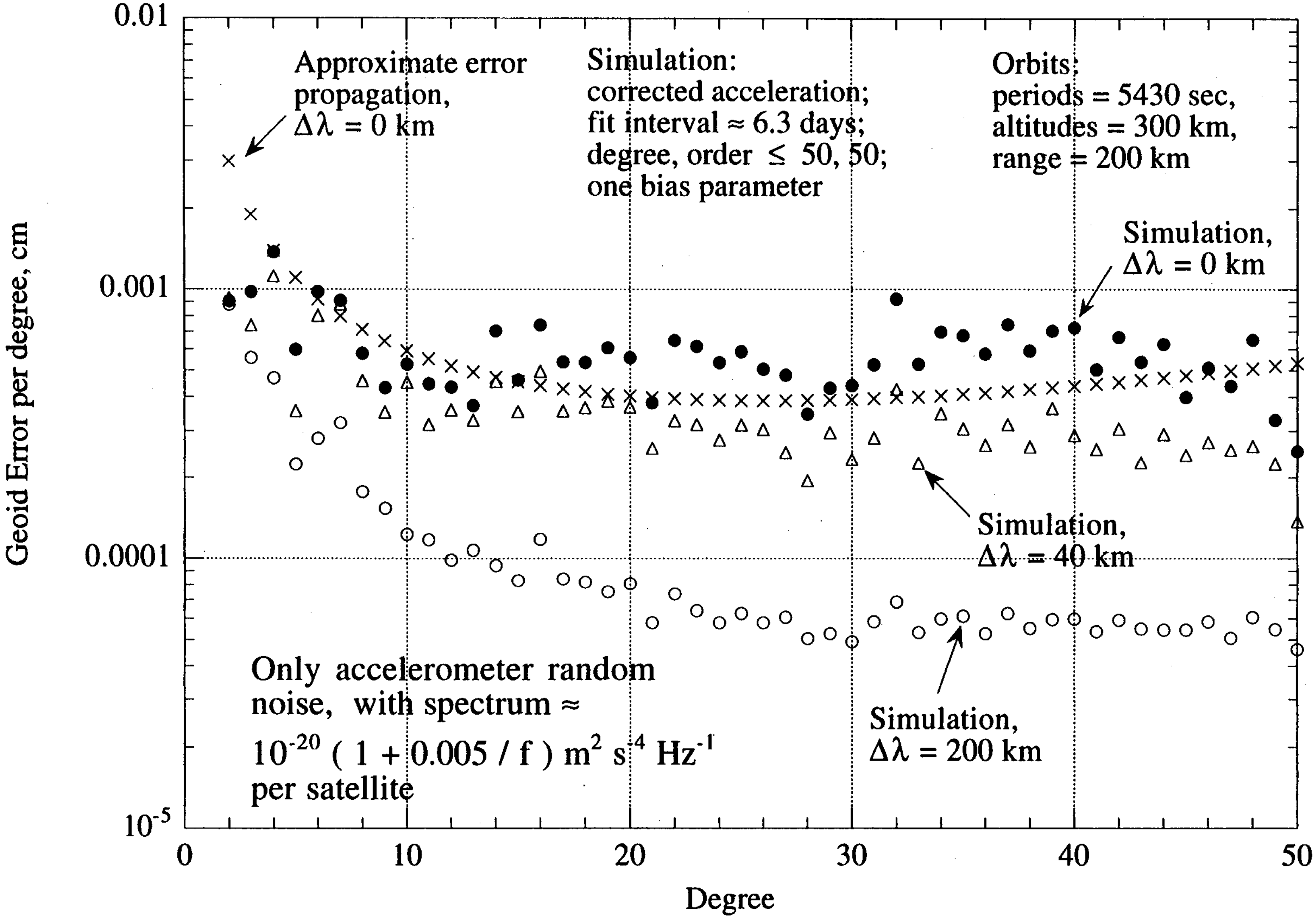

Figure 6-3 Example Simulations of Geoid Error Caused by Accelerometer Random Error in Corrected Acceleration Observable

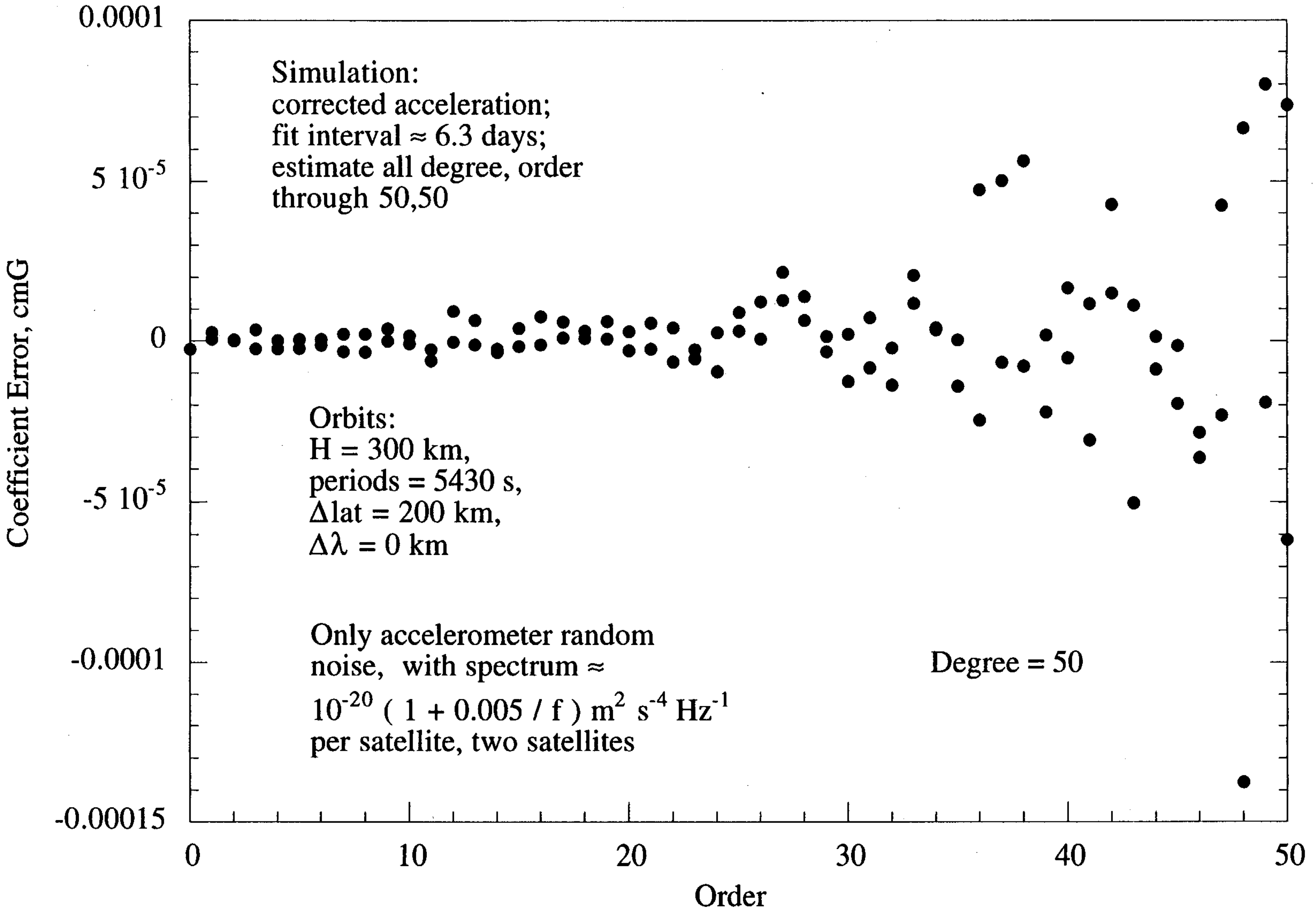

Figure 6-4 Example Simulation of Gravity-Coefficient Errors Caused by Accelerometer Noise when Longitude Separation is Zero

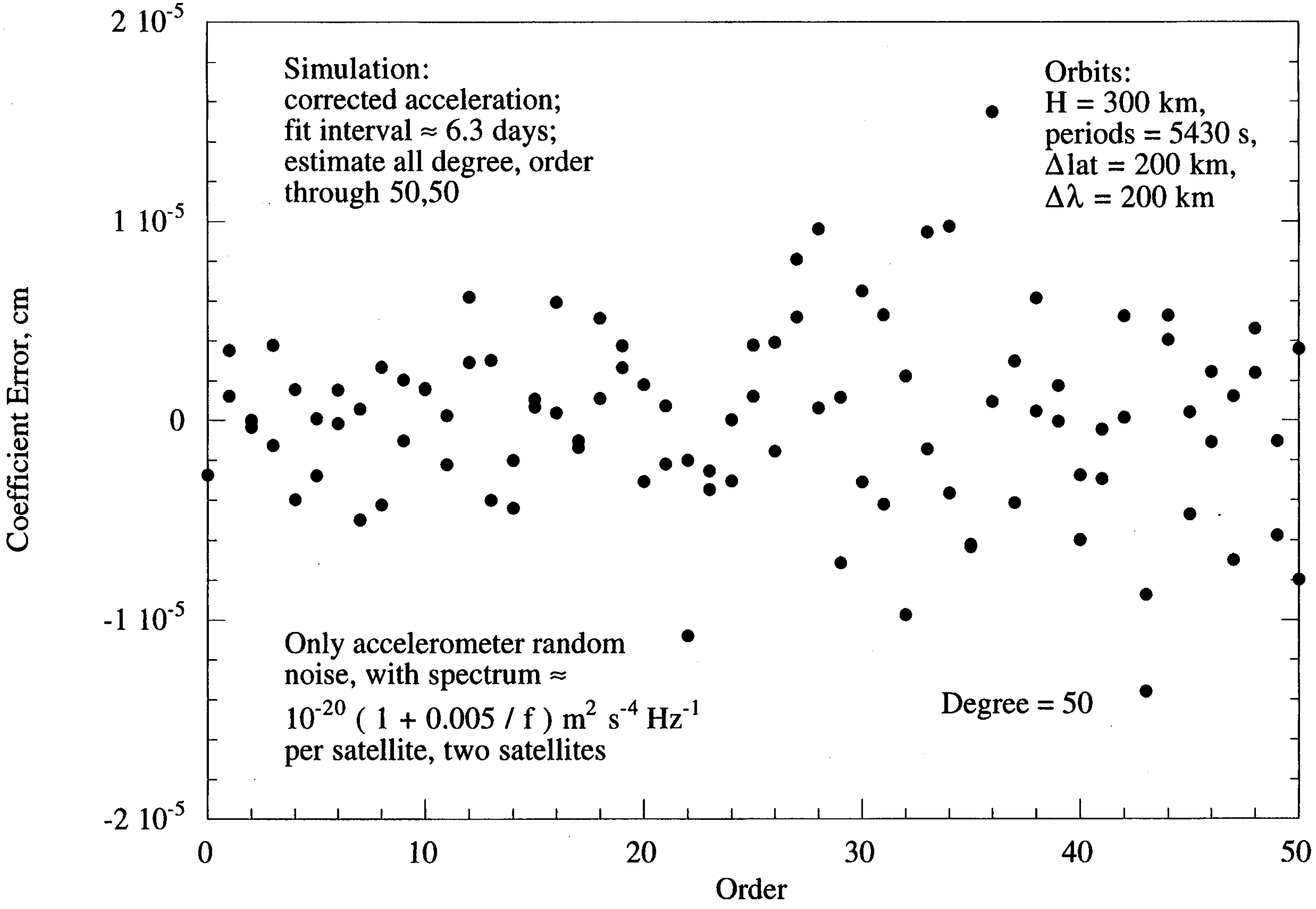

Figure 6-5 Example Simulation of Gravity-Coefficient Errors Caused by Accelerometer Noise when Longitude Separation is 200 km

APPENDIX A

EXTRACTION OF RANGE, RANGE RATE AND ACCELERATION

As explained in Section 2, dual-1-way (biased) range values are to be produced at a nominal rate of 10 S/s. A method is needed to extract (biased) range, range-rate and acceleration observables at a reduced rate (e.g., 0.1 S/s) from these "high-rate" range values. A critical feature to be investigated is the noise filtering in the frequency domain caused by the method used to generate these observables. As discussed in Section 5, estimation of a given harmonic coefficient can be viewed as an operation that extracts from the observables the overall amplitude of a sum of tone frequencies that characterize that particular harmonic. Thus, the error in the estimation of a given harmonic coefficient will be determined by the noise at the specific tone frequencies for that harmonic. Since the estimation process is so frequency specific, error analysis should also be carried out in the frequency domain, with a careful accounting of all noise mapped into the band of frequencies (the "signal band") covered by harmonic tones. For gravity coefficients up to 100x100, the signal band for K-band observables extends from approximately 0.1 mHz to 18 mHz. For gravity coefficients up to 200x200, the signal band extends from approximately 0.1 mHz to 36 mHz. Since gravity coefficients are amplitude quantities, another important feature of the observable-extraction method is tone amplitude distortion (gain ripple) introduced in the signal band by processing. Ripple specifications are discussed in more detail below.

This appendix first analyzes noise mapping associated with a standard quadratic least-squares fit and shows that the quadratic-fit approach is suboptimal for extracting range, very suboptimal for range rate, and extremely suboptimal for range acceleration. Other "fitting" approaches are then considered and shown to be suboptimal. A digital-filter approach is then presented for extracting these three observables, an approach that provides better performance with regard to noise mapping and adequately preserves gravity harmonic amplitudes across the signal band.

The particular class of finite-impulse-response (FIR) digital filters analyzed in this appendix is not submitted as the best class for processing GRACE range data but is introduced, because of its closed-form simplicity in the frequency domain, to illustrate concepts and demonstrate feasibility. More work is needed to determine the best filters for GRACE, particularly with regard to minimizing filter time span.

A.1 Quadratic-Fit Approach

A.1.1 Quadratic-Fit Equations

One possible method for extracting the range, range-rate and acceleration observables from the 10-S/s range values is the well-known technique of fitting a quadratic function by least squares to the range values to extract constant (range), linear (range rate) and quadratic (acceleration) parameters. Such a fit could be applied sequentially to successive 10-s intervals to produce the three desired observables at a reduced rate of, say, 0.1 S/s.

A least-squares fit of a quadratic function to range values within a given 10-s interval can be represented by the well-known expression:

$$\mathbf{P} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{R}' \qquad \text{(A.1)}$$

where **P** is the vector representing the three estimated parameters (observables):

$$\mathbf{P} = \begin{pmatrix} R \\ \dot{R} \\ \ddot{R} \end{pmatrix} \qquad \text{(A.2)}$$

R' is the vector representing the N range values (e.g., N = 100) in the given interval, defined as

$$\mathbf{R}' = \begin{pmatrix} R'_1 \\ R'_2 \\ R'_3 \\ \cdot \\ \cdot \\ R'_N \end{pmatrix} \qquad \text{(A.3)}$$

and $\mathbf{A}^T$ is the transpose of the partials matrix given by

$$\mathbf{A}^T = \begin{pmatrix} 1 & 1 & 1 \;\ldots\ldots & 1 \\ t_1 & t_2 & t_3 \;\ldots\ldots & t_N \\ \frac{1}{2}t_1^2 & \frac{1}{2}t_2^2 & \frac{1}{2}t_3^2 \;\ldots\ldots & \frac{1}{2}t_N^2 \end{pmatrix} \qquad \text{(A.4)}$$

where t_k is the time tag of range value k. Time is referenced to interval center so that t_k is zero at midinterval and so that extracted observables will have a time tag at interval center.

A.1.2 Frequency Response of Quadratic Fit

One can determine the frequency response ("fit filter") generated by a quadratic fit for each of the three observable types by setting the input range values in Eq. (A.1) equal to an appropriate sinusoid. Because of symmetry considerations, the input range is set equal to $\cos(2\pi f t_k)$ for the estimated range and acceleration parameters and to $\sin(2\pi f t_k)$ for the range-rate parameter. (The quadrature sinusoid in each case produces zero response as a result of a symmetry considerations.) This approach leads to the following approximate expressions for fit filters in terms of "voltage" response:

$$\sqrt{G_R} \approx -1.5\left[\frac{\sin(\pi f T_f)}{\pi f T_f} + 5\frac{\cos(\pi f T_f)}{(\pi f T_f)^2} - 5\frac{\sin(\pi f T_f)}{(\pi f T_f)^3}\right] \qquad \text{(A.5)}$$

for range,

$$\sqrt{G_{\dot{R}}} \approx -\frac{6}{T_f}\left[\frac{\cos(\pi f T_f)}{\pi f T_f} - \frac{\sin(\pi f T_f)}{(\pi f T_f)^2}\right] \qquad \text{(A.6)}$$

for range rate, and

$$\sqrt{G_{\ddot{R}}} \approx \frac{60}{T_f^2}\left[\frac{\sin(\pi f T_f)}{\pi f T_f} + 3\frac{\cos(\pi f T_f)}{(\pi f T_f)^2} - 3\frac{\sin(\pi f T_f)}{(\pi f T_f)^3}\right] \qquad (A.7)$$

for acceleration, where the nominal value for the fit interval, T_f, is 10 s and f is the "test" frequency.

The fit filters in Eqs. (A.5), (A.6) and (A.7) are plotted as function of frequency in Figs. A-1, A-2, and A-3 in terms of both "voltage" and "power" responses. In order to make the plots independent of fit interval T_f, all three cases are plotted as a function of normalized frequency, $f T_f$, and the response amplitudes for range rate and acceleration are plotted in "unnormalized" form, with normalization explained in the plot comments. For GRACE, the plots can be easily converted to actual values, for example, by substituting the nominal fit interval of $T_f = 10$ s. The figures show only the responses for positive frequencies since the responses for negative frequencies are either symmetric or antisymmetric replicas. Based on $T_f = 10$ s, all three plots illustrate the frequency range ("signal band") of the gravity signal for degree up to 100.

As indicated in Figs. A-2 and A-3 for small frequencies, the filter simply multiplies a range component at frequency f by $2\pi f$ for range rate and by $-(2\pi f)^2$ for acceleration, respectively, as one would expect for a differentiating operation. Thus, for signal-band frequencies, both the signal and the noise components in the range-rate and acceleration observables are greatly reduced in magnitude at lower frequencies relative to components at higher frequencies. The "resonance" maxima in the range-rate and acceleration filters at approximately $f = 0.7/T_f$ and $1.1/T_f$, respectively, occur at the frequency values for which sinusoid time dependence across the fit interval is closest to linear and quadratic form, respectively. These maxima, in combination with the aforementioned relatively lower filter gain at low frequencies, turn out to be particularly damaging as explained in the next subsection.

The frequency response for the range observable in Fig. A-1 is not unlike the sinx/x response associated with a straight average, but is somewhat wider. Fractional amplitude distortion (gain ripple) is greatest at the indicated upper edge (degree = 100) of the signal band, reaching approximately –0.00055 (–65 dB) at $fT_f = 0.2$ ($f = 0.02$ Hz with $T_f = 10$ s). (For $T_f = 5$ s, amplitude distortion has the same value since fT_f is still 0.2 at the upper edge of the signal band, given maximum degree = 200 with a maximum frequency of approximately 0.04 Hz.) As discussed below, amplitude distortion of this magnitude is marginally acceptable, but can be reduced with better processing. For the range-rate filter in Fig. A-2, fractional amplitude distortion (relative to $2\pi f$) is considerably worse, reaching the unacceptable level of –0.039 (–28 dB) at $fT_f = 0.2$. For the acceleration filter in Fig. A-3, corresponding fractional amplitude distortion (relative to $(2\pi f)^2$) is –0.028 , which is also too large.

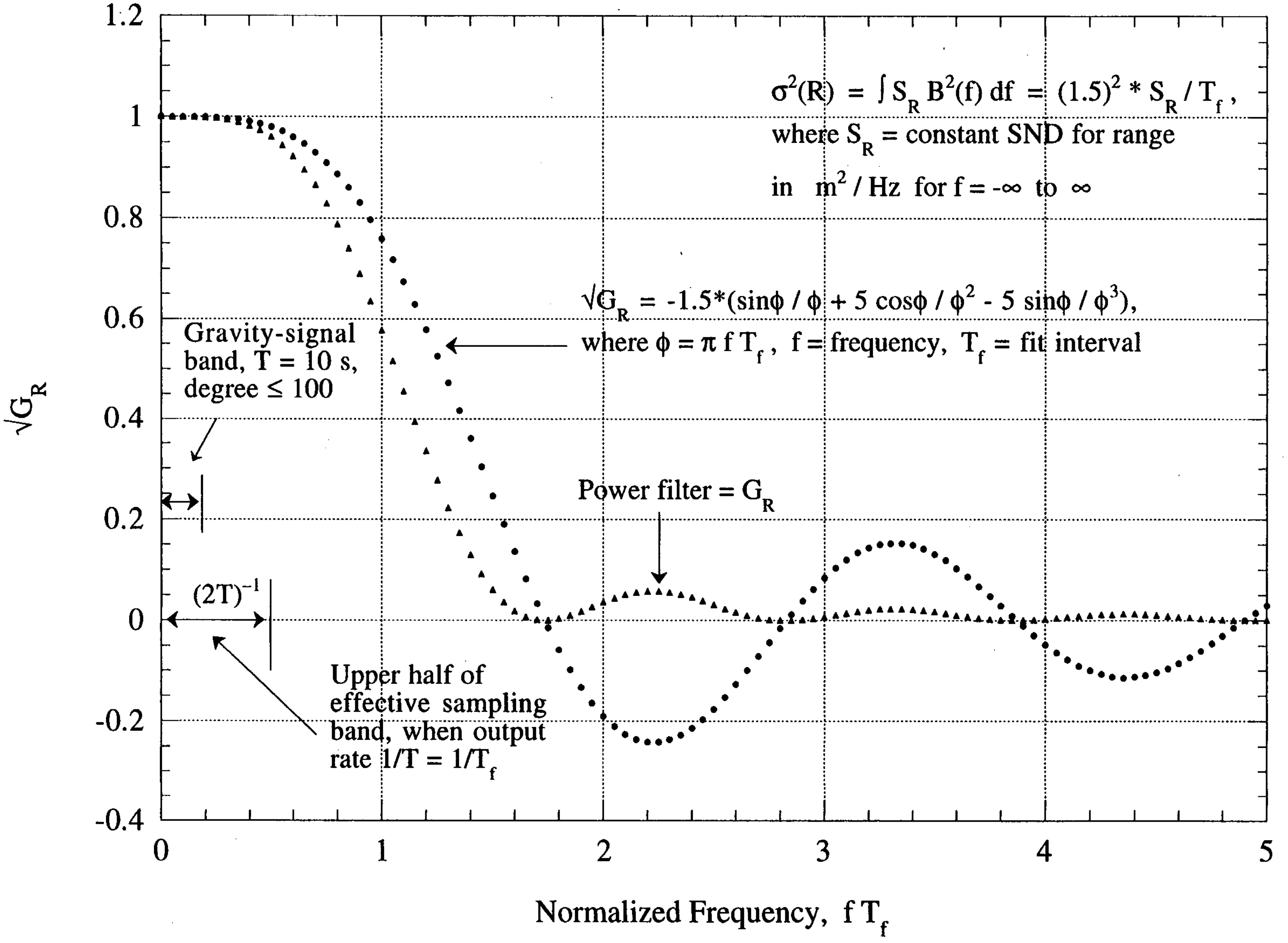

Figure A-1 Fit Filter for Range Observables Obtained from a Quadratic Fit to Range Values

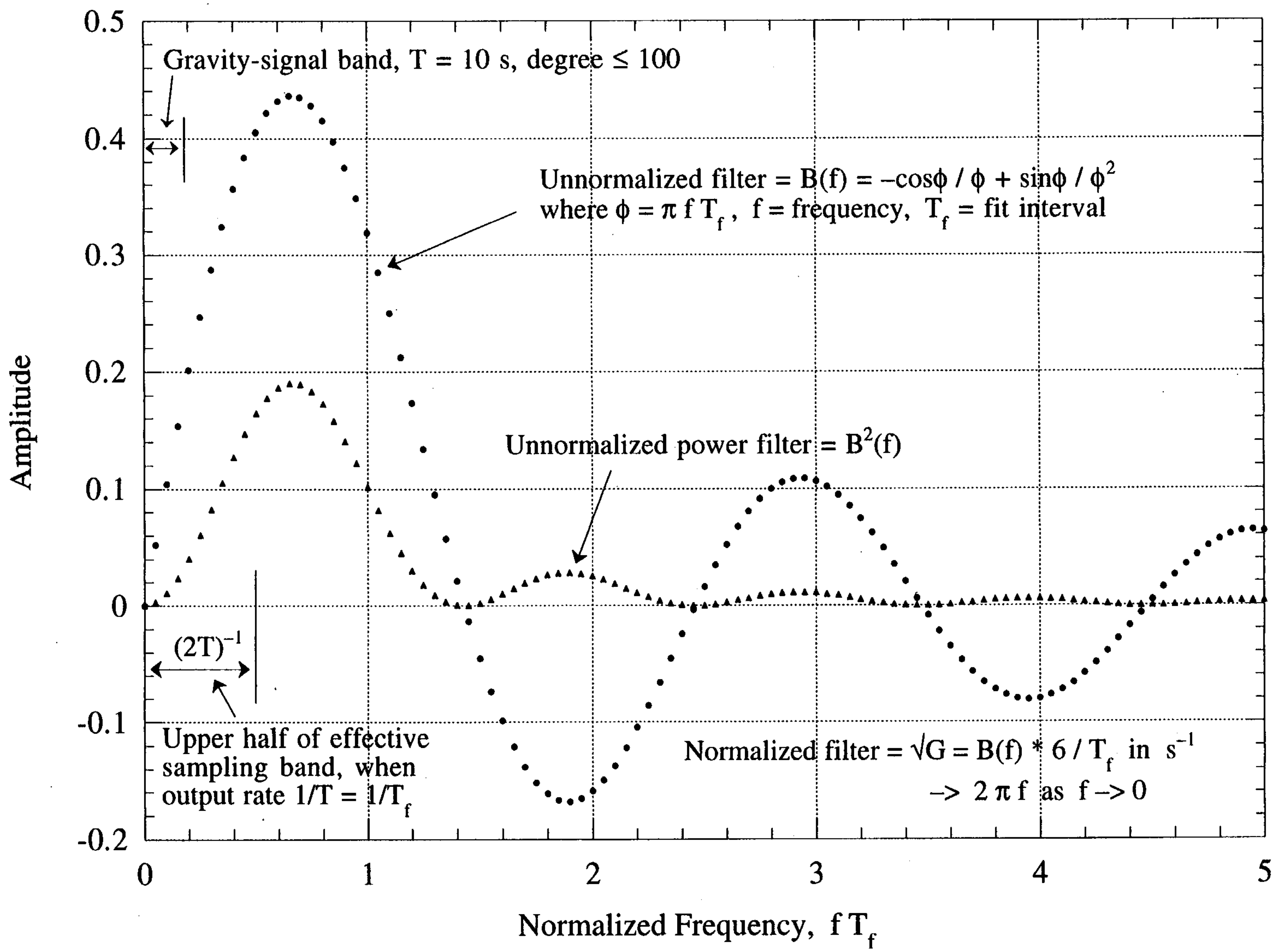

Figure A-2 Fit Filter for Range Rate Obtained from a Quadratic Fit to Range Values

92

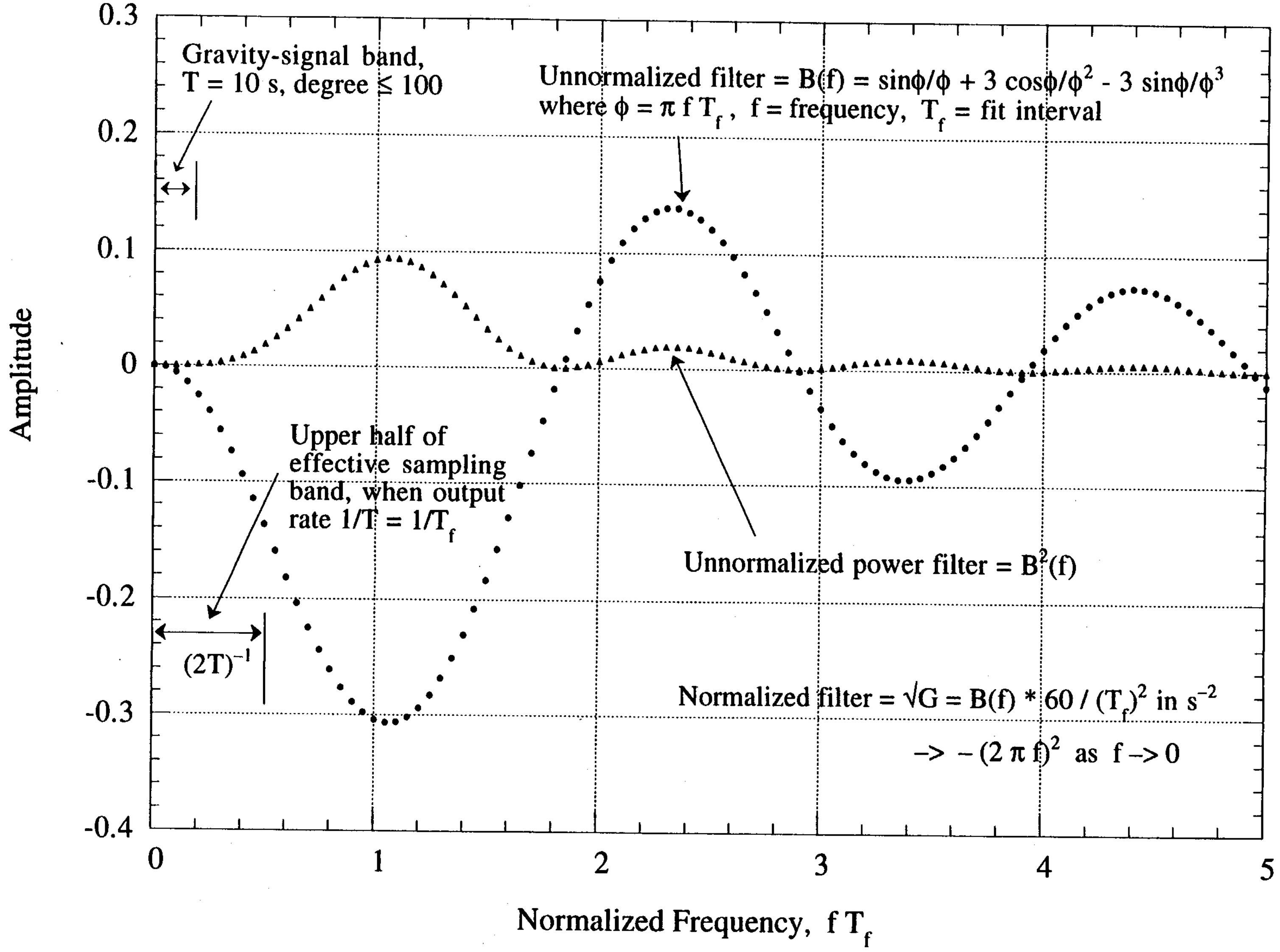

Figure A-3 Fit Filter for Acceleration Observables Obtained from a Quadratic Fit to Range Values

A.1.3 Sampling Passband and Aliasing

If one assumes that range, range rate and acceleration are to be generated at a sample rate of 1/T, the effective "sampling passband" can be viewed as 0 Hz $\pm 1/(2T)$. Any frequency component falling outside this range is "aliased" into the sampling passband as a result of the sampling operation. (Only the positive-frequency half of the filter is illustrated in Figs. A-1, A-2, and A-3.) That is, aliasing shifts any given "out-of-band" frequency component into the sampling passband by "modulo'ing" said frequency with 1/T (i.e., by shifting it by $\pm n/T$ where the integer n and the sign are selected to place the resulting frequency in the sampling passband). Thus, if the processing filter (frequency response) to obtain the observables does not suppress noise at frequencies outside the sampling passband, that noise will alias into the passband, thereby increasing effective noise on targeted signal components in the passband. In ideal processing, the effective filter for the process would eliminate frequency components outside of the sampling passband (i.e., for $|f| > 1/(2T)$).

Assume at this point that time is divided into adjacent, nonoverlapping intervals of length T and that a separate quadratic fit is applied to each interval to extract one observable per interval. In this case, the sample rate is equal to the inverse of the fit interval: $1/T = 1/T_f$. As indicated by Figs. A-1, A-2, and A-3, the fit filters resulting from a quadratic fit with this fit interval fall short of meeting the out-band-filtering requirement for all three observable types. For all three observables, significant power is passed by the fit filter at frequencies with magnitudes greater than $1/(2T)$. Though not illustrated in these figures (see next subsection), that out-of-sampling-band power aliases into the indicated sampling passband to degrade the signal band. Thus, this quadratic fit approach is clearly a suboptimal way of extracting GRACE observables.

A measure of the overall impact of aliasing in the case of range can be obtained by applying the quadratic-fit filter to the spectral noise density for input biased range, S_R, and integrating across all frequencies to obtain range noise variance:

$$\sigma_R^2 = \int_{-\infty}^{\infty} S_R(f)\, G_R(f)\, df \tag{A.8}$$

where G_R is the range fit filter given in Fig. A-1. For constant ("white") spectral noise density, one can show this integral yields

$$\sigma_R = 1.5 \sqrt{S_R \left(\frac{1}{T_f}\right)} \tag{A.9}$$

This result indicates that range noise is a factor 1.5 greater for a quadratic fit than that for ideal filtering ($\sqrt{S_R/T}$), given an output sample rate of $1/T = 1/T_f$ when the quadratic fit interval is T_f. That is, for white input noise, better filtering can improve total effective noise on range by a factor of 1.5 relative to the assumed quadratic-fit.

A.1.4 Effective Frequency Response for White Noise After Aliasing

To gain more detailed insight, the effective increase in range error due to out-of-band noise can be computed at <u>each</u> signal-band frequency across the sampling passband by aliasing all out-of-band noise into the sampling passband and root-sum-squaring the noise at each frequency. It is again assumed that the sample rate (1/T) is equal to the inverse of the quadratic-fit interval (T_f). Results for the special case of <u>white noise</u> are presented in Fig. A-4. The figure, which is now scaled for 10-s sample points, also shows the unaliased part of the fit filter below 0.1 Hz, as displayed at low frequencies in Fig. A-1. Note that the effective noise increase at each frequency in the sampling passband is a constant factor of 1.5 relative to ideal filtering (with a gain of 1.0) across the sampling passband. This result is consistent with the total integrated noise computed above.

The suboptimality of the quadratic-fit method is more extreme for range rate and even more extreme for acceleration. For both of these observables, the large "resonance" maxima that fall outside of the sampling passband (see Figs. A-2 and A-3) will alias into the passband. This aliasing can be particularly damaging for these observables since the amplitudes of signal and noise at low signal-band frequencies is reduced to relatively low levels as a result of differentiation. Thus, because of the amplitude disparity between aliased out-of-band noise and signal-band noise, aliasing can cause a dramatic decrease in signal-to-noise ratio (SNR) at given signal-band frequencies.

This effect is illustrated in Figs. A-5 and A-6 for range rate and acceleration for the special case of *white noise* on the biased range values input to the fit. In both figures, noise has been aliased, as described for range above, into the sampling passband based on the fit filters given by Eqs. (A-6) and (A-7). For both observable types, out-of-band noise that is aliased into the sampling passband results in a *constant* effective noise "gain" across the sampling passband, as in the case of range in Fig. A-4. The effects of the amplitude disparity between out-of-band noise and in-band noise, mentioned in the last paragraph, can now be quantified for each signal-band frequency. For example, at the low signal-band frequency of approximately 0.4 mHz (2/rev), a quadratic-fit filter increases effective noise by approximately a factor of 150 for range rate, and by a factor greater than 10^4 for acceleration, relative to an ideal filter that eliminates out-of-band noise. (The unaliased part of the quadratic-fit filter shown in the figures approximates an ideal filter in these comparisons.) These results pertain to white noise on the input biased-range values; results for other noise spectra can be dramatically different. For example, the increase in effective noise at 2/rev can be much worse in the case of input "blue" noise (i.e., with relatively large high-frequency components), but can be much smaller for "red" noise (i.e., with relatively small high-frequency components).

A possible modification for reducing the negative impact of aliased noise in quadratic fits is to increase the quadratic-fit interval from 10 s to, say, 40 s, but keeping the output rate at the nominal value of 0.1 S/s. This strategy would improve noise errors in range rate and acceleration by keeping the resonance peaks out of the signal band but it would push amplitude distortion in the signal band to even more unacceptable levels, as suggested by the fall off in the range filter amplitude in Fig. A-1.

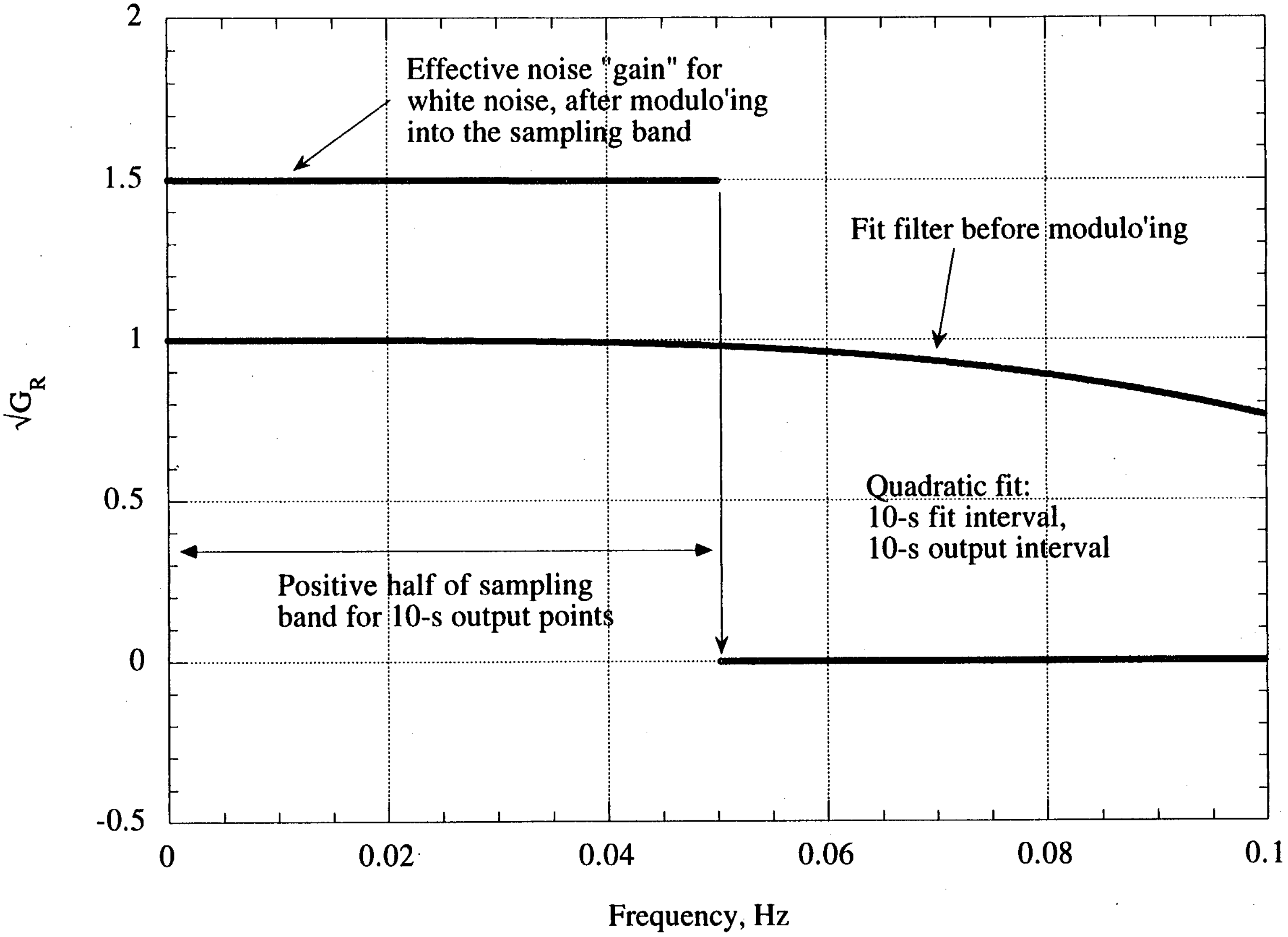

Figure A-4 Effective Filter Applied to White Noise by a Quadratic Fit to Obtain Range

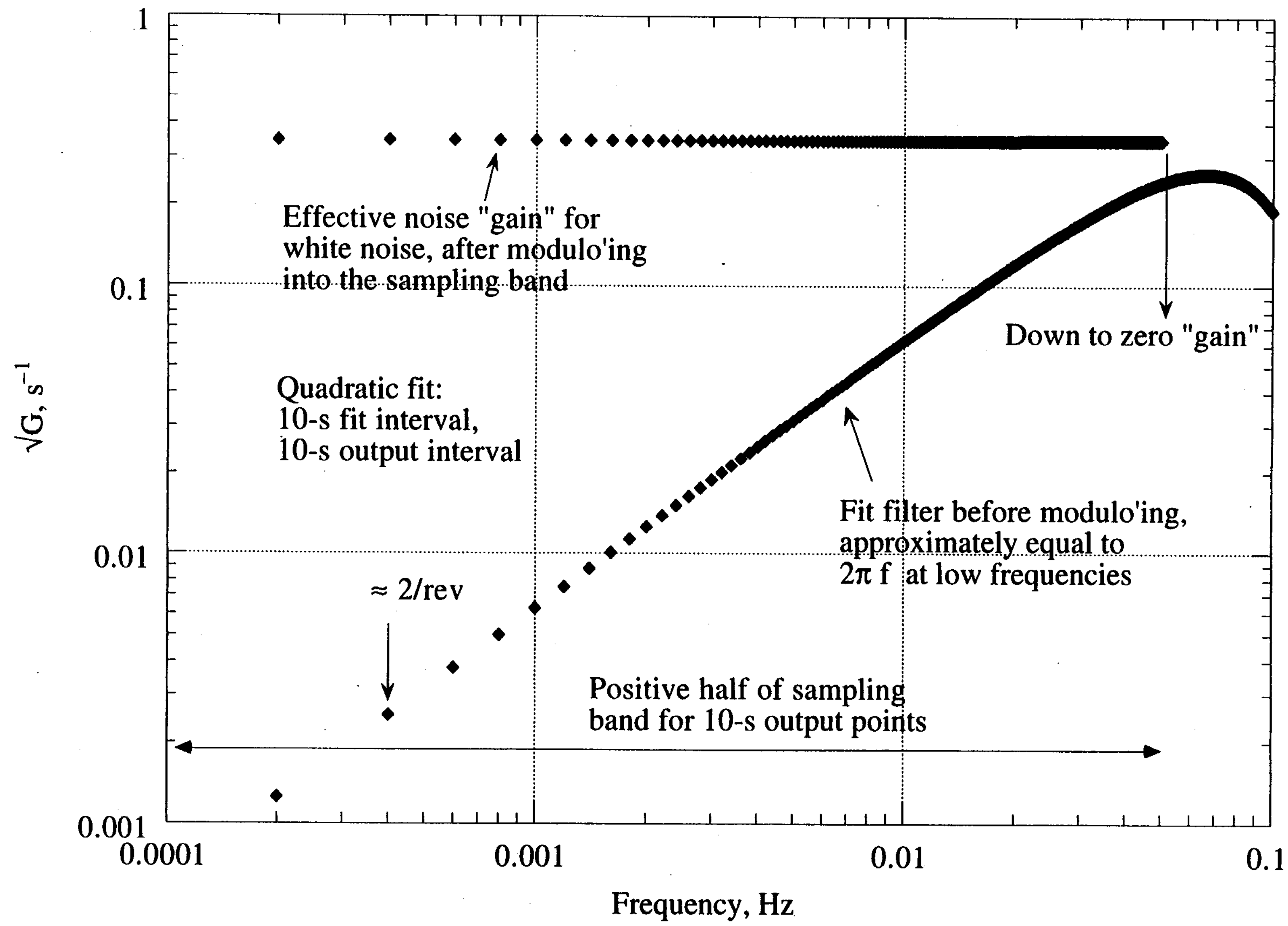

Figure A-5 Effective Filter Applied to White Noise by a Quadratic Fit to Obtain Range Rate

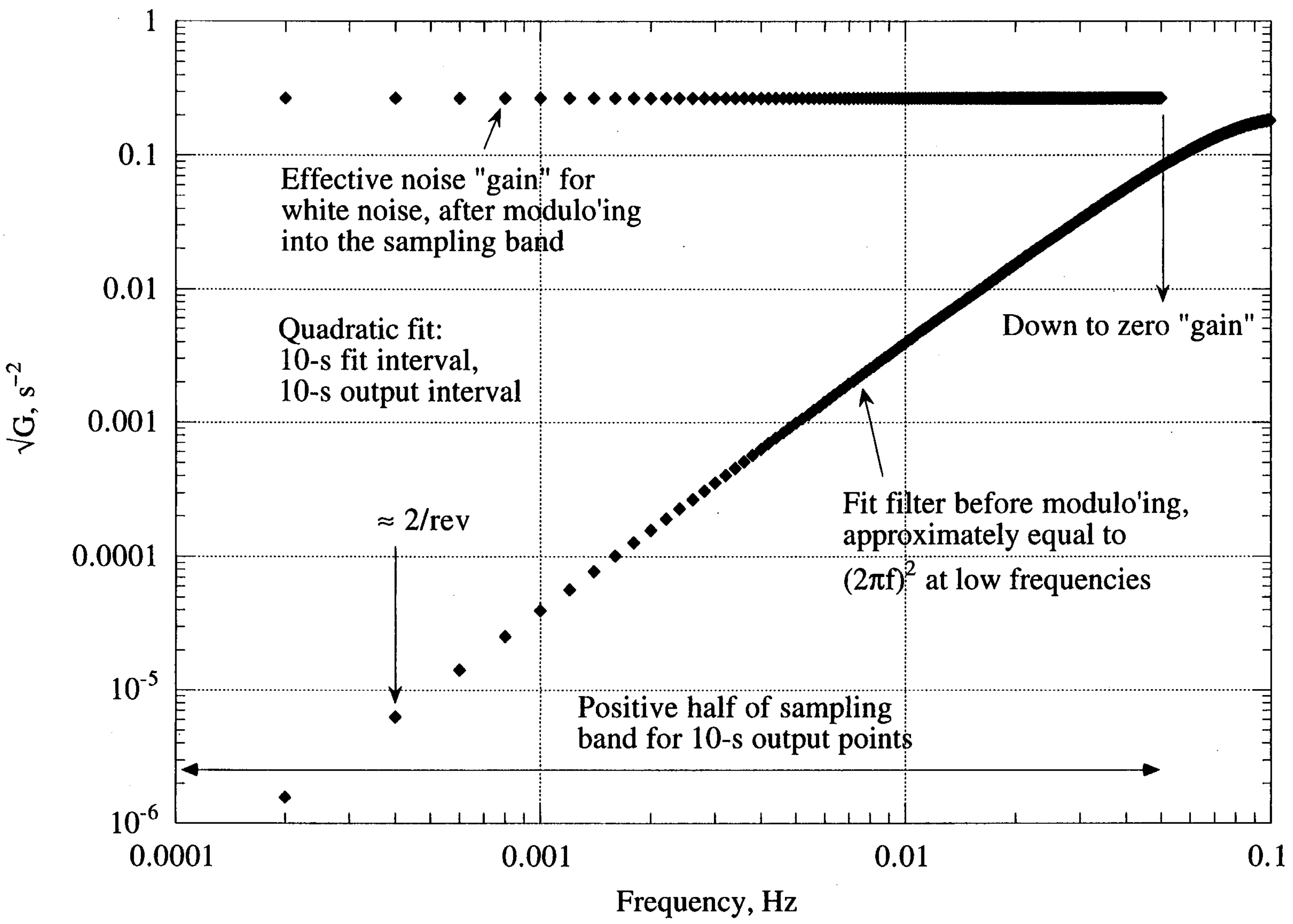

Figure A-6 Effective Filter Applied to White Noise by a Quadratic Fit to Obtain Acceleration

A.2 Other Fitting Approaches

An alternative to a quadratic fit is to fit a linear function $(R + \dot{R}t)$ over the fit interval T_f to estimate range and range rate at interval center. One can easily show that the resulting frequency-domain filter for range is approximately equal to $\sin(\pi f T_f)/(\pi f T_f)$. With respect to noise mapping in the case $1/T = 1/T_f$, this range filter is somewhat better than the corresponding filter (see Fig. A-1) from a quadratic fit (approximately a factor of 1.5 better for white noise), but it causes much worse amplitude distortion at the upper edge of the signal band (e.g., fractional amplitude distortion is –0.065 at $fT_f = 0.2$ vs. –0.00055 for the quadratic fit). The frequency-domain filter for range rate resulting from a linear fit is the same as the unacceptable range-rate filter from a quadratic fit, namely the filter presented in Fig. A-2. Thus, except for the noise improvement in range, a linear-fit approach suffers from the same unacceptable shortcomings (amplitude distortion and/or noise aliasing) as a quadratic fit.

Another approach for extracting range rate is presented in Reference [1]. Since that reference does not explicitly treat amplitude distortion as an error or present range-rate spectral errors in terms of $\mu m\ s^{-1}\ Hz^{-1/2}$ as a function of frequency across the signal band, after aliasing, it is difficult to directly compare the results obtained in that paper with the present analysis. The following analysis attempts to cast the approach of that paper in the same terms as the current analysis. In the "triangular weighting" option of Reference [1], range is effectively "differentiated" by averaging range separately over two adjacent intervals, each of length T', and then differencing the average values and dividing by T'. Expressed in terms of the analysis of this report, these operations result in a frequency-domain filter operating on range, with the form, $[\sin(\pi f T')/(\pi f T')]^2\, 2\pi f$. Reference [1] does not explicitly relate the averaging length T' to an output rate for range rate (denoted as 1/T here, equal to, say, 1/10 s in analogy with the preceding fits). First assume the length of each adjacent averaging interval is half the interval between output points, $T' = T/2$. (For example, for 10-s averaging intervals, the output rate would be 1 sample/20-s. This assumption appeals to some because it keeps input samples from overlapping between output range-rate points.) In this case, the fractional amplitude distortion (i.e., $1 - [\sin(\pi f T')/(\pi f T')]^2$) at the upper edge of the signal band is comparable to the unacceptable linear-fit filter (e.g., –0.032 at $fT' = 0.1$ vs. –0.039 for a linear fit over T) and aliased noise is very large and also comparable to the unacceptable linear-fit filter (larger than a linear fit with $T = T_f$ by approximately a factor of 1.16 for white noise). Decreasing the length of the output interval so that $T' = T$ (i.e., 10-s averaging and an output rate of 1 sample/10-s), reduces aliased noise to a low level but makes unacceptable amplitude distortion even worse (e.g., –0.125 at $fT' = 0.2$). These considerations indicate the range-rate approach of Reference [1] has unacceptable shortcomings similar to those for the linear- and quadratic-fit filters discussed above, namely unacceptable aliased noise and/or amplitude distortion in the signal band.

Because the results in Reference [1] are presented for each value of averaging-interval length in terms of RMS noise, care must be taken in interpretation. RMS noise is not a sufficiently informative way of representing noise and can give a misleading impression, particularly when aliasing is considered, when the spectral noise density is far from uniform, or when the selected averaging length is not appropriate for the targeted gravity-signal band (e.g, for a maximum gravity-signal frequency of 0.02 Hz, an averaging interval of 1 s is so short that noise outside the gravity-signal band is inappropriately counted and 100 s is so long that noise is not counted over the whole signal band). In assessing a dual-1-way technique, the important quantities to consider are spectral noise density and amplitude distortion, each as a function of frequency across the targeted gravity-signal band, after accounting for all processing steps, including aliasing. (Gravity-signal band and output rate are determined by the highest gravity harmonic to be estimated, as discussed in Subsection 3.1.5.) The two choices for output rate discussed above for triangular weighting, namely $T = T'$ and $T = 2T'$, provide a good example of how RMS noise can be misleading. For an averaging length of T' = 10 s, which would be a possible choice for a maximum signal-band frequency of, say, 0.01 Hz, the output-sample spacings of T = 10 s and T = 20 s would have the same RMS noise but the T = 20 s case would have far worse noise in the signal band as a result of aliasing.

The noise-magnifying and/or amplitude-distortion shortcomings of these techniques make it clear that a better method for extracting the intersatellite observables is needed, particularly for range rate and acceleration. Given the sampling passband imposed by the observable output rate of 1/T, the extraction method should apply a low-pass filter with very flat signal-band response and a relatively sharp cut-off at 1/(2T) so that aliasing is reduced to negligible levels. The rest of this appendix develops digital filters that provide much better performance. (Under a general definition of digital filter, the preceding approaches can each be cast as a digital filter for each observable. However, in accord with convention, the more common labels are used here.)

A.3 Digital-Filter Approach

A.3.1 Definitions

The following paragraphs develop finite-impulse-response (FIR) digital filters for extracting range, range rate and acceleration from the dual 1-way range values by simultaneously applying a low-pass filter, reducing the data rate, and, for range rate and acceleration, applying the appropriate differentiation. For illustration here, the filter output rate is assigned a nominal value 0.2 S/s. (With straightforward digital-filter changes and reprocessing, the output rate can readily be changed to accommodate different goals.) The three time-domain filters, denoted by F_n, $\dot{F}_n$, $\ddot{F}_n$, operate in the time domain on the measured range values (R'_i at time point i sampled at a nominal input sample rate of $f_s = 10$ S/s) to generate the observables of range, range rate, and range acceleration, respectively, at the nominal output rate of 1 sample per 5 s (0.2 S/s):

$$R_i = \sum_{n=-N_h}^{N_h} F_n \, R'_{i-n} \tag{A.10}$$

$$\dot{R}_i = \sum_{n=-N_h}^{N_h} \dot{F}_n \, R'_{i-n} \tag{A.11}$$

$$\ddot{R}_i = \sum_{n=-N_h}^{N_h} \ddot{F}_n \, R'_{i-n} \tag{A.12}$$

where $N_h = (N_f - 1) / 2$. N_f is the total number of points spanned by the filter and is given by $N_f = f_s T_f$, where T_f is the filter time span. (For computational convenience, N_f will be set equal to an odd number.) The index i denotes the i^{th} output value computed as a weighted sum of the N_f input points symmetrically surrounding the i^{th} output point in time. For each new output point, the index i increments by the number of input time points equivalent to the desired output interval, which nominally corresponds to 10 S/s * 5 s = 50 points for an output interval of 5 s.

A.3.2 Digital-Filter Specifications

The digital filters must differentiate to the desired order and have a nearly rectangular response in the frequency domain with an appropriate low-pass cutoff. To avoid aliasing out-of-band noise into the signal band, the low-pass cutoff is set approximately equal to the Nyquist value, which is 0.1 Hz for an output sample rate of 0.2 S/s. This cutoff would provide, as far as the digital filter is concerned, ample margin for gravity harmonics up to degree 200, with a maximum tone frequency of 0.036 Hz. (To estimate harmonics up to degree 100, a lower output rate could be used, as discussed in Section 3.) The frequency response must have sufficiently small sidelobes and cause amplitude distortion (e.g., gain ripple) across the signal band (0.1 to 36 mHz in this example) at a level that satisfies the project goal of 0.01 cmG per degree for gravity coefficients.

Since gravity harmonics of lower degree generally have higher amplitude, the gain-ripple requirement in terms of fractional deviation is more demanding for the lower-degree harmonics, which are close to the passband center (i.e., zero frequency). The most demanding ripple requirement applies to the J2 term, which has an amplitude on the order of 3 kmG. For this unique harmonic, the gain ripple requirement is very stringent, say, 0.01 cmG / 3 kmG = 3x 10^{-8} = –150 dB at a frequency of approximately 0.37 mHz. Thus, the ripple requirement is very demanding near the passband center, but it can be satisfied with careful filter design. The next largest harmonics, which are close to J2 in frequency, are on the order of 300 times smaller, and therefore can be assigned a far less demanding ripple requirement on the order of –100 dB. As a result of J2 proximity, this requirement is easily satisfied if the J2 requirement is met.

A crude requirement for the other, higher-degree coefficients can be derived as follows. Kaula's rule predicts that the sum of the squares of all coefficients associated with a given degree value (n), K_n^2, varies as a function of degree according to the approximate expression:

$$K_n^2 \approx \frac{1.6\text{x}10^{-10}}{n^3} \tag{A.13}$$

in dimensionless units, which can be rewritten in terms of cm of equivalent geoid error as

$$K_n \approx \frac{8000\ \text{cmG}}{n^{1.5}} \tag{A.14}$$

Thus, the ripple requirement becomes less stringent as degree, and therefore harmonic frequency, increases. If one adopts a nominal per-degree error of 0.01 cmG for gravity coefficients and assumes the 2n+1 coefficients for degree n have roughly the same magnitude but random signs, one can show that a crude gain-ripple specification is given by

$$\varepsilon < \frac{0.01\ \text{cmG}}{K_n} = \frac{n^{1.5}}{800{,}000} \tag{A.15}$$

where ε, the fractional coefficient error caused by gain ripple, is assumed to be the same for all coefficients for that degree value. For degree 100, Kaula's rule predicts $K_n \approx 8$ cm so the ripple requirement prescribed by Eq. (A.15) is $\varepsilon = 0.0013$ or –58 dB at a nominal frequency of 18 mHz.

The specification in Eq. (A.15) is assumed to apply to the ripple at a bandpass frequency equal to the maximum tone frequency for degree n. That is, in effect, the specification is based on the assumption that all harmonics of a given degree have an effective frequency equal to said

100

maximum tone frequency. This assumption makes this ripple requirement very conservative. Since the effective frequencies of harmonics of high order for a given degree are much lower than the maximum harmonic frequency for that degree value, as discussed in Section 5, the high-order harmonics actually have effective frequencies that fall at lower points in the passband where gain ripple tends to be smaller, as shown below. Thus, since the derivation of Eq. (A-15) is based on the assumption that all harmonics experience the same gain ripple as the maximum frequency, Eq. (A.15) is a conservative requirement for ripple error. A more rigorous derivation of ripple requirements needs to be carried out.

A requirement on frequency-domain sidelobes is needed to limit aliasing of out-of-band noise into the signal band, but it is less demanding than the J2 ripple requirement. A requirement of, say, –70 dB or less for maximum sidelobe magnitude is adequate for GRACE applications.

A.3.3 The Windowing Approach for Digital Filters

A digital filter with a perfect rectangular response in the frequency domain would be a suitable filter for processing GRACE data. As is well known, such a filter can be theoretically generated by convolving a signal with a sinx/x function in the time domain, provided the range of the sinx/x function is extended to infinity. Practical constraints, however, require the sinx/x function to be limited to a time window of tractable length. Such truncation can be implemented as multiplication by a rectangular window function. For example, the digital filter in Eq. (A.10) for time point n can be specified as

$$F_n = W_n \frac{\sin(2\pi B n/f_s)}{2\pi B n/f_s} \tag{A.16}$$

where W_n is the rectangular window function, $1/f_s$ is the time between input sample points, and B is the single-sided bandwidth that would be produced by a sinx/x function of infinite extent. The window function is set to zero outside the desired filter time span (e.g., a rectangle given by $W_n = 1.0$ for $|n| \le N_h$ and $W_n = 0$ for $|n| > N_h$). Because of the sharp transition at the window edges, a rectangular window function causes defects in the resulting frequency-domain filter, including gain distortion across the band, a skirt on the frequent cutoff and sidelobes.

To mitigate these problems, a well-known approach (e.g., see [6]) is to multiply the sinx/x function by a graduated window function (e.g., the time-domain windows of Hamming, von Hann, Blackman, and Kaiser) that slowly decreases to zero at the edges of the window. That is, the graduated window function smoothes the transition from sinx/x dependence to zero weight, thereby reducing ripple and sidelobes. Even with graduated windowing, however, the frequency response of the filter will still be corrupted by sidelobes, gain distortion and cutoff skirt, but to a lesser extent than for a rectangular window function.

For GRACE, the gain-distortion requirement, in particular for gain-ripple near the center of the passband, is particularly demanding. Gain ripple is improved by constructing a window function whose Fourier transform has a fast sidelobe drop-off rate. This relation can be readily explained on the basis of the convolution theorem. That is, the product of the sinx/x function and the window function in the time domain in Eq. (A.16) becomes a convolution of their respective Fourier transforms in the frequency domain. The Fourier transform of the sinx/x is, of course, the desired rectangular response in the frequency domain. The convolved transform deviates from this rectangular shape as a result of the finite width of the mainlobe and the finite sidelobes of the window transform, with the window-mainlobe width causing most of the width of the filter cut-off skirt of the convolved transform and the window sidelobes causing both the filter gain ripple within the passband and the filter sidelobes outside the passband. Thus, it is the shape of the transform of the window function (i.e., the width of its main lobe and the magnitude and drop-off rate of its sidelobes) that determines the deviation in overall filter response from the desired rectangular shape. Because of their finite length, all window functions cause some deviation from rectangular shape, so the task is to select a window function that meets the requirements of the application.

A.3.4 The CRN Class of FIR Digital Filters

This report does not make a quantitative comparison of the many available digital filter designs to determine the best design for GRACE. The CRN class of FIR digital filters developed in the following subsections is not submitted as the optimal class for GRACE processing but is presented, because of closed-form simplicity in the frequency domain, to illustrate concepts and demonstrate feasibility. This class of filters provides transparent flexibility in adjusting frequency-domain characteristics over a wide range, provides more-than-adequate sidelobes and gain ripple, allows simple modification to provide differentiating capability, and leads to specific candidate filters that satisfy the nominal GRACE requirements. The Blackman and Kaiser windows, which are not analyzed here, are also promising candidates for meeting the GRACE requirements. More work is needed to determine which window can provide the best performance with regard to gain ripple and filter time span.

The approach used to develop the CRN class of filters deviates from the "time-domain-first" windows referenced above in that the filter, for both computational simplicity and interpretive convenience, is first constructed in the frequency domain and then the time-domain form is obtained by Fourier transforming to the time domain. An advantage of this class of filters is that the mainlobe width and sidelobe amplitudes of the window transform can be easily hand-calculated in advance on the basis of a simple closed-form expression. Differentiation can be easily added to an otherwise satisfactory low-pass design by simply multiplying the selected frequency-domain shape by $-2\pi f$ for the first derivative and by $-(2\pi f)^2$ for the second derivative before transforming to the time domain. Similarly, somewhat arbitrary shapes can be applied to the frequency response by simply multiplying by the appropriate function in the frequency domain. This last capability is very useful for generating simulated random noise with selected shapes in frequency domain.

With the "frequency-domain-first" approach, care must be taken in crafting the initial desired frequency-domain response to make certain the resulting time-domain filter actually ends up having a finite window (i.e., is exactly zero outside of the specified time span). This requirement and easy flexibility are both made possible by the class of time-domain window functions formed by convolving a rectangular time-domain window with itself N_c times. This class of digital filters will be referred to as CRN filters, denoting N self-convolutions of a rectangular time-domain window function. Since the Fourier transform of a rectangular window in time has a sinx/x form in the frequency domain, the convolution theorem shows that the frequency-domain response of an N-fold self-convolved time-domain rectangle is $(\sin x/x)^N$. Thus, the frequency-domain form of a CRN window function is a very simple expression that is easy to compute in parametric fashion and easy to assess in terms of its own mainlobe and sidelobes, and consequently in terms of sidelobes, gain ripple and skirt width of the ultimate digital filter.

The CRN-filter class offers a transparent means for easily generating a range of options for the mainlobe width and sidelobe magnitude of the window-function transform. As the number of self convolutions (N_c) increases, the sidelobes of the Fourier transform of the window function decrease rapidly in the form of a small number to the N_c power. For $N_c = 5$, for example, the first and second sidelobes are reduced by a ("voltage") factor approximately equal to $(2/3\pi)^{N_c}$ (i.e., –67 dB in power) and $(2/5\pi)^{N_c}$ (–90 dB in power), respectively, relative to the main lobe. As mentioned above, such rapid sidelobe drop-off in the transform of the window function greatly helps in reducing gain ripple in the ultimate filter frequency response. Increasing the number of convolutions from 5 to 7 decreases the magnitude of the first sidelobe by approximately a factor of 22 (by –27 dB to –94 dB) so that increasing the number of convolutions is a very effective way to decrease gain ripple and sidelobe amplitude.

A negative consequence of increasing the number of convolutions is a result of the fact that N_c self convolutions of a basis rectangle of length T_R produces a filter with a time span, T_f, given approximately by N_cT_R. That is, for a selected total window time span T_f, the width of the basis rectangle decreases in inverse proportion to N_c and therefore the width of the main lobe of the window transform, and the spacing of its sidelobes, increase as N_c. Consequently, the skirt width of the low-pass cut-off of the ultimate filter broadens and its sidelobes and ripple spread out as N_c increases for fixed T_f. The spreading of sidelobes means sidelobe magnitude does not drop off as rapidly as a function of frequency due to sidelobe count, but this is more than compensated for by the power of N_c effect on sidelobe magnitude, as explained above. For GRACE, limited broadening of skirt width is acceptable, up to the point that gain distortion spreads back into the passband to effect the highest required signal frequencies. As mentioned above, the most demanding ripple requirement is near zero frequency where skirt broadening has little influence.

Frequency-domain examples of CR3, CR5, CR7 and CR9 window functions are presented in Fig. A-7, each computed on the basis of a filter time span of $T_f = 100$ s. In this frequency-domain plot (in "voltage"), the absolute value of the sidelobes has been plotted so that a log scale could be used. That is, the odd sidelobes are actually negative. The corresponding time-domain window functions are shown in Fig. A-8. As one would expect, in progressing from CR3 to CR9, the time-domain window functions taper off ever more gradually from center to edge, while the mainlobes in the frequency domain increase in width and the sidelobes spread out and decrease in magnitude.

A.3.5 Filter Design Tradeoffs

In designing a CRN digital filter, one has to choose values for bandwidth, the number of convolutions N_c, and filter length T_f. Bandwidth should be set to the Nyquist value of half the output rate of the filter, or less (e.g., ≤ 0.1 Hz for a 0.2 S/s output rate). Larger values of N_c offer the advantage of making sidelobe magnitudes smaller in the window-function transform. However, as N_c increases for a fixed filter length, the width of the mainlobe of the window-function transform increases, which broadens the skirt of the low-pass cut-off of the digital filter so that it can reach back into the signal band and distort gain when the mainlobe becomes too wide. For GRACE, the signal band is in the middle part of the sampling band, which makes the skirt distortion less troublesome. (For example, given an upper limit on the signal band of 36 mHz for degree ≤ 200, the highest signal frequency is a comfortable 64% below a cut-off frequency 0.1 Hz. Further, the gain-distortion requirement at 36 mHz of –49 dB, based on the crude specification approach described above, is far less stringent than the distortion requirement of –150 dB at 0.36 mHz.) As filter length increases for fixed N_c, gain distortion at a given signal-band frequency tends to decrease as a result of the narrowing of the mainlobe of the window transform and the more rapid ripple drop-off caused by closer ripple spacing. Thus, increasing filter length is advantageous with respect to gain distortion, but has limits due to practical considerations such as computation time and exacerbation of data gaps. In the design below, filter length is adjusted to a relatively low value consistent with signal-band-gain-distortion goals.

Because of the sign flips of sidelobes found in odd-N_c CRN filters, a step up from even to odd N_c provides a much greater reduction in ripple than a step up from odd to even. For this

103

reason, only odd N_c filters are considered ($N_c = 3, 5, 7$ and 9). A complete analysis has not been carried out at this time to determine the best filter parameters, but crude tradeoffs based on the aforementioned considerations suggests that a reasonably satisfactory filter could be constructed on the basis of $N_c = 5$ or 7. In order to completely illustrate a given set of filter parameters, the detailed examples below assume the following nominal parameters: 0.2 S/s output rate, approximately 0.1 Hz bandwidth, $T_f \approx 100$ s and $N_c = 7$. Many more options for filters can be obtained by varying parameters, as illustrated for a few cases in Subappendix A.3.8. More analysis is needed to determine the "optimal" filters, including more rigorous treatment of ripple specifications and of filter time span.

A.3.6 An Example of a CRN Digital Filter for the Range Observable

To generate amplitude values for the selected CRN digital filter in the time domain, first create the required frequency response in the frequency domain and then transform with a discrete Fourier transform (DFT) to the time domain. (A strict DFT approach in all aspects of this process has been selected for achieving the demanding specifications for gain distortion.) Let $N_f = f_s{*}T_f = 2N_h + 1$ be the odd number of time points spanned by the desired digital filter of length T_f in time, where f_s is the sample rate of the input points (e.g., 10 S/s). Let the DFT of the digital filter have N_f frequency points with a spacing of $1/T_f$ and a frequency span of approximately $-f_s / 2$ to $+f_s / 2$. If B is the target single-sided bandwidth of the desired low-pass filter, let $N_B \approx B\,T_f$ be the number of frequency points spanned by B and for convenience let $2N_B + 1$ be an odd number. Note that the integer form of N_B means that B is approximated in quantized steps with a step size of $1/T_f$ (e.g., 0.01 Hz for $T_f = 100$ s). Let N_R be equal to the number of points spanned by the basis rectangle of length T_R, where $T_R = N_R/f_s$. For computational simplicity, select N_R and N_f so that $N_f = N_c N_R$ where N_c, is the number of convolutions of the basis rectangle. For a CR7 filter with an output rate of 0.2 S/s, a bandwidth of 0.1 Hz and a time span of approximately 100 s, example values are: $N_f \approx (10 \text{ S/s}){*}100\text{s} \rightarrow 1001$ (i.e., $T_f = 100.1$ s); $N_R = 1001/7 = 143$ (i.e., $T_R = 14.3$ s) and $N_B \approx 0.1\text{-Hz}{*}100.1 \text{ s} \rightarrow 10$.

As described above, the desired frequency-domain response (i.e., DFT of the desired time-domain digital filter) is constructed as a convolution of a rectangular low-pass filter with the DFT of the selected time-domain window function. As discussed above, the rectangular low-pass filter has a quantized single-sided bandwidth equal to N_B/T_f and an amplitude equal to 1.0. Thus, with the quantization conventions specified above, the unnormalized frequency response can be computed according to

$$H_k = \sum_{k'=-N_B}^{k'=N_B} \left(\frac{\sin[\pi(k-k')/N_c]}{\sin[\pi(k-k')/N_f]} \right)^{N_c} \qquad \text{(A.17)}$$

where the convolution sum over k' is carried out over points within the desired bandpass ($\pm N_B/T_f \approx \pm B$), and H_k is the resulting frequency response at frequency point k/T_f, $|k| \leq N_h$. The $(\sin x/\sin y)^N$ function in Eq. (A.17) is the DFT of the selected CRN time-domain window function that is formed from a N_c-fold self-convolution of a rectangular function of width $N_R = N_f/N_c$. (The sinx/x function alluded to above actually becomes the sinx/siny function in Eq. (A.17) when a DFT rather than a Fourier transform is involved.)

Once the frequency response has been computed according to Eq. (A.17), the unnormalized time-domain digital filter for the range observable can be generated with a final DFT:

$$F_n = \sum_{k=-N_h}^{N_h} H_k \cos(2\pi k n/N_f) \quad \text{for } |n| \le N_h \qquad \text{(A.18)}$$

An overall gain normalization must be applied to the filter. This gain normalization is determined by applying (as in Eq. A.10) the time-domain filter to a range input, R'_{i-n}, formed by a cosine function with i = 0 and with unit amplitude:

$$R'_{-n} = \cos[2\pi f n / f_s], \quad |n| \le N_h \qquad \text{(A.19)}$$

in which the frequency f is set approximately equal to 2/rev, the dominant tone in J2. For these normalization calculations, the 2/rev frequency has been assigned a nominal value of 2/5400 s ≈ 0.37 mHz. (As discussed above, the J2 harmonic is selected for normalization since J2 sets the most demanding ripple requirement.) Gain normalization (i.e., the factor multiplying F_n) is adjusted so that the amplitude of this particular cosine function is still 1.0 (within round off) after being passed through the filter. For the filter parameters discussed above, this normalization process provides a filter that passes the amplitude of a 0.37-mHz tone with a fractional gain error (roundoff error) of only 5×10^{-15} or –286 dB, which far exceeds the requirement of –150 dB cited above. (In this approach, fractional gain error increases significantly relative to –286 dB for frequency points slightly displaced from 0.37 mHz, but the gain-ripple specifications are still satisfied, as illustrated below.)

Fig. A-9 presents a plot of the resulting range digital filter, F_n. As one would expect, the filter resembles a sinx/x function in the center portion but tapers to zero toward the edges. The window function causing this tapering, which is presented in Fig. A-8, is equivalent to a 7-fold self-convolution of a rectangular function with a width of T_R = 100.1 s / 7 = 14.3 s. Actual frequency response of the filter can be directly and accurately tested by applying the filter to a set of unit-amplitude cosine functions, given by Eq. (A.19), but with frequency, f, now stepped to span the band of interest. The resulting frequency response, which is presented in Fig. A-10, shows a cutoff at approximately 0.1 Hz as required, with acceptable sidelobe amplitudes (approximately –100 dB vs. the nominal requirement of –70 dB).

Gain ripple relative to an ideal response of 1.0 is presented in Fig. A-11 along with a curve representing maximum allowed ripple as approximately derived from Kaula's rule according to Eq. (A.15) and along with other options for filter time span (see Subappendix A.3.8). As can be seen, actual ripple for T_f = 100.1 s surpasses the approximate Kaula-based requirement across the allowed signal band (nominally 0.1 to 50 mHz) by approximately two orders of magnitude (i.e., contributing 1 µm or less to geoid error per degree). This signal band would accommodate harmonics up to degree 270. Note the decrease in ripple error to approximately 10^{-9} (–180 dB) at 0.4 mHz, which is near the adopted normalization reference point of 0.37 mHz with a ripple specification of –150 dB.

Similar plots can be generated for an equivalent CR7 digital filter with a 0.05-Hz bandwidth that would produce output points at the rate of 0.1 S/s. The parameters for such a filter would be $N_f \approx$ (10 S/s)*200 s –> 2009 (i.e., T_f = 200.9 s); N_R = 2009/7 = 287 (i.e., T_R = 28.7 s) and $N_B \approx$ 0.05-Hz*200.9 s –> 10. Figs. A-7 through A-15 would be nearly identical for such a filter except that the time span of the filter would increase from 100 s to 200 s and the frequency axis, where relevant, would scale by a factor of two. Thus, such a filter could accommodate harmonics up to degree 135 (up to 25 mHz), with gain ripple error on the order of 1 µm or less.

106

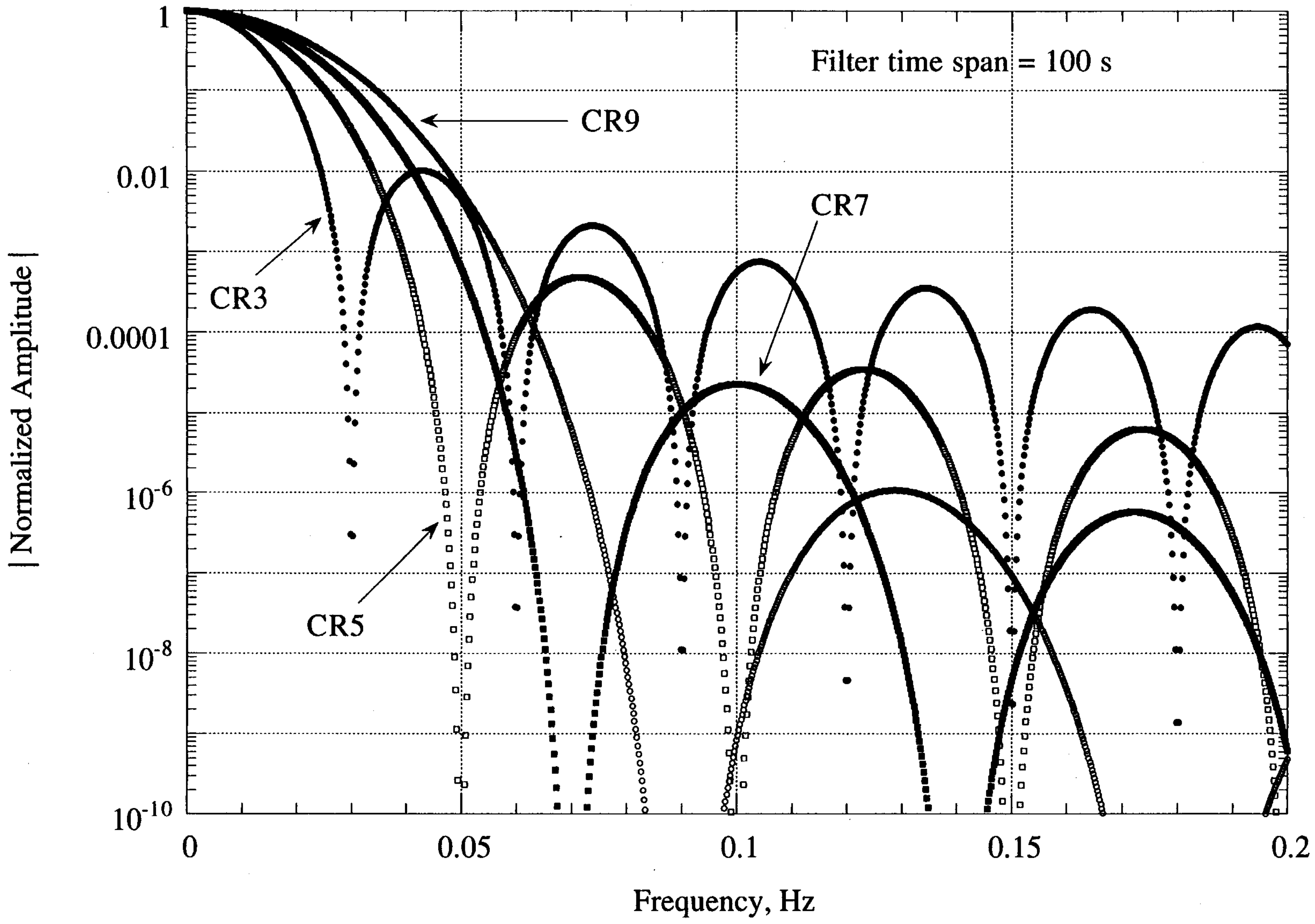

Figure A-7. Fourier Tranforms of the Window Functions for Four CRN Digital Filters

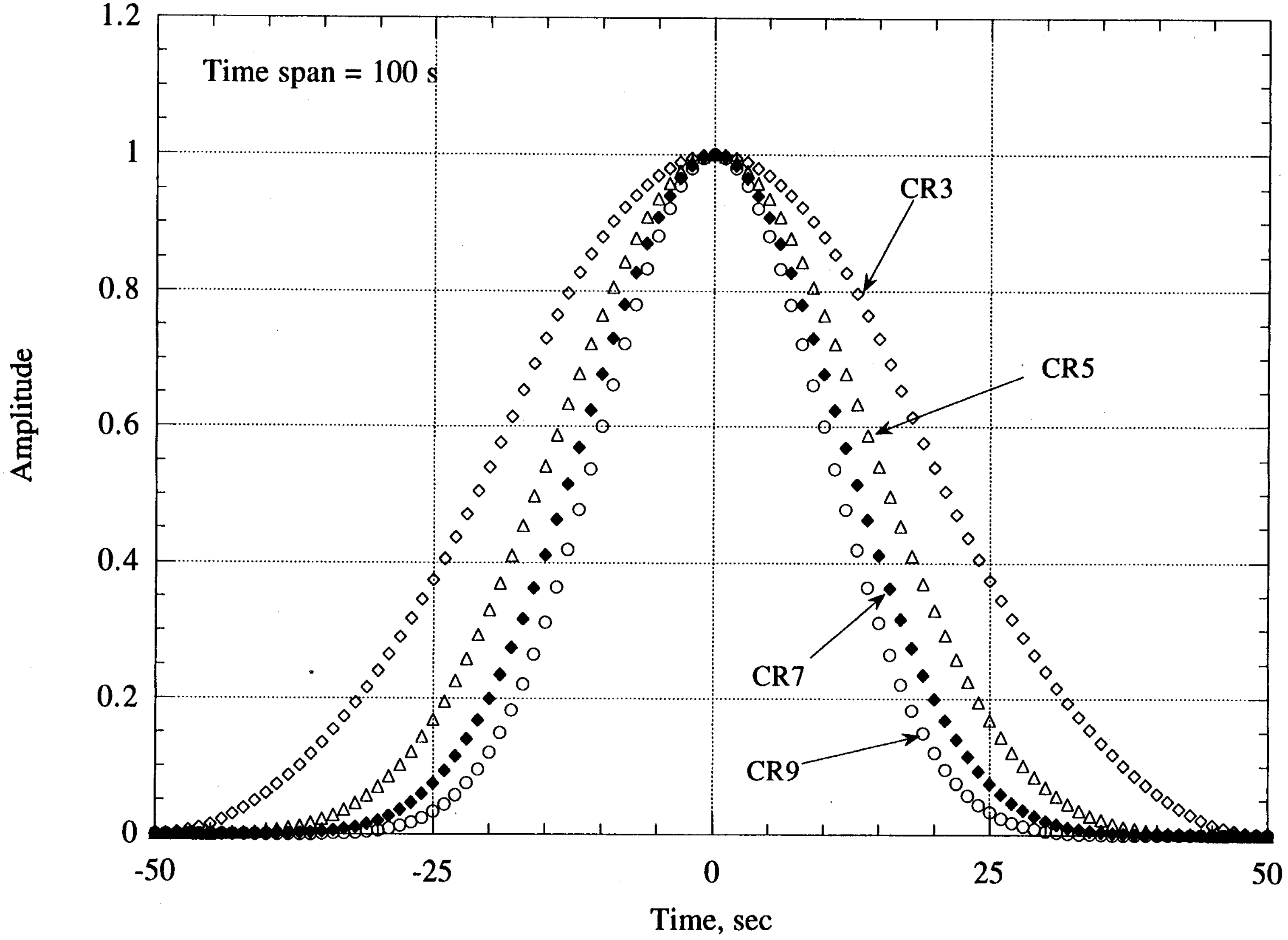

Figure A-8 Window Functions for Selected CRN Digital Filters

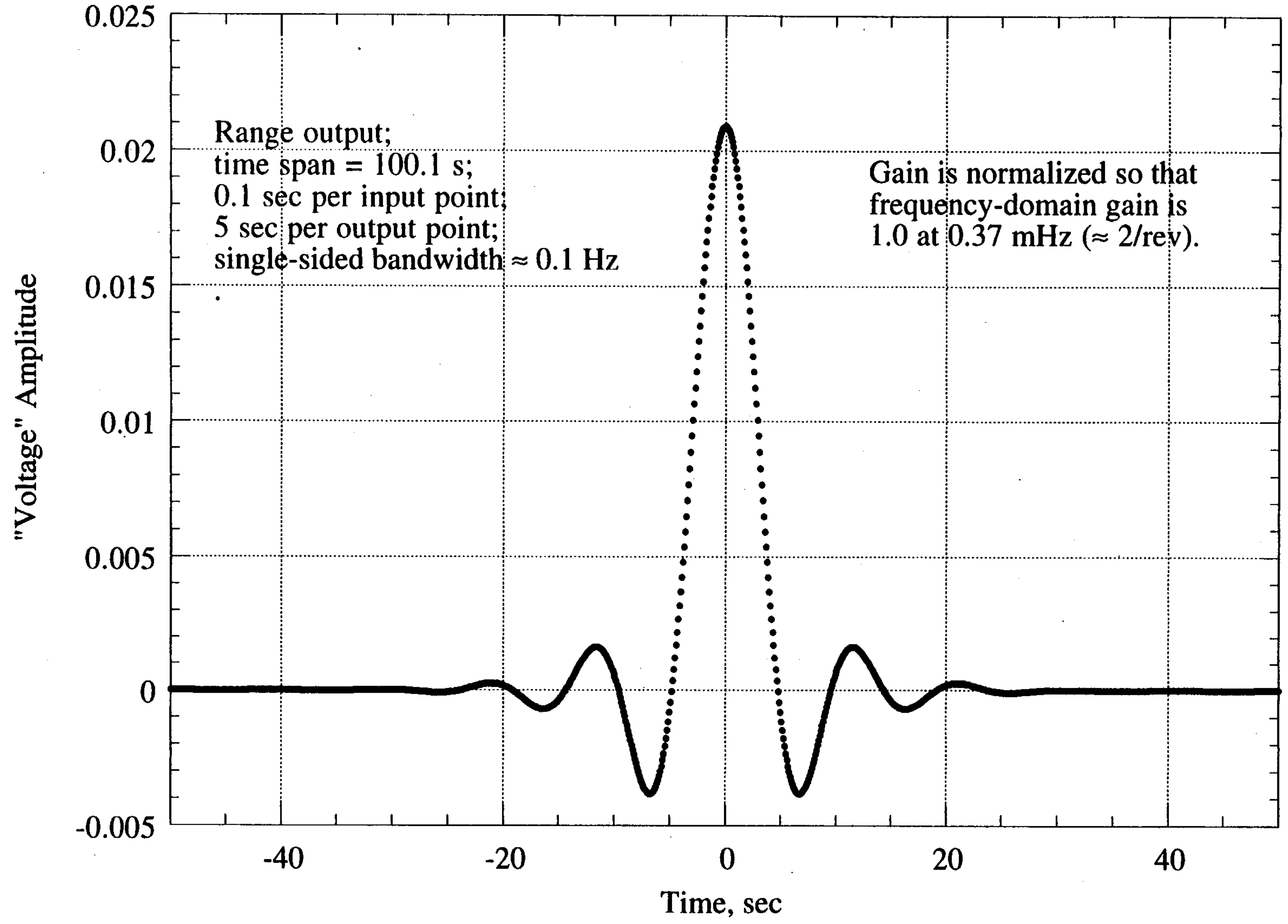

Figure A-9 An Example of a CR7 Digital Filter in the Time Domain

109

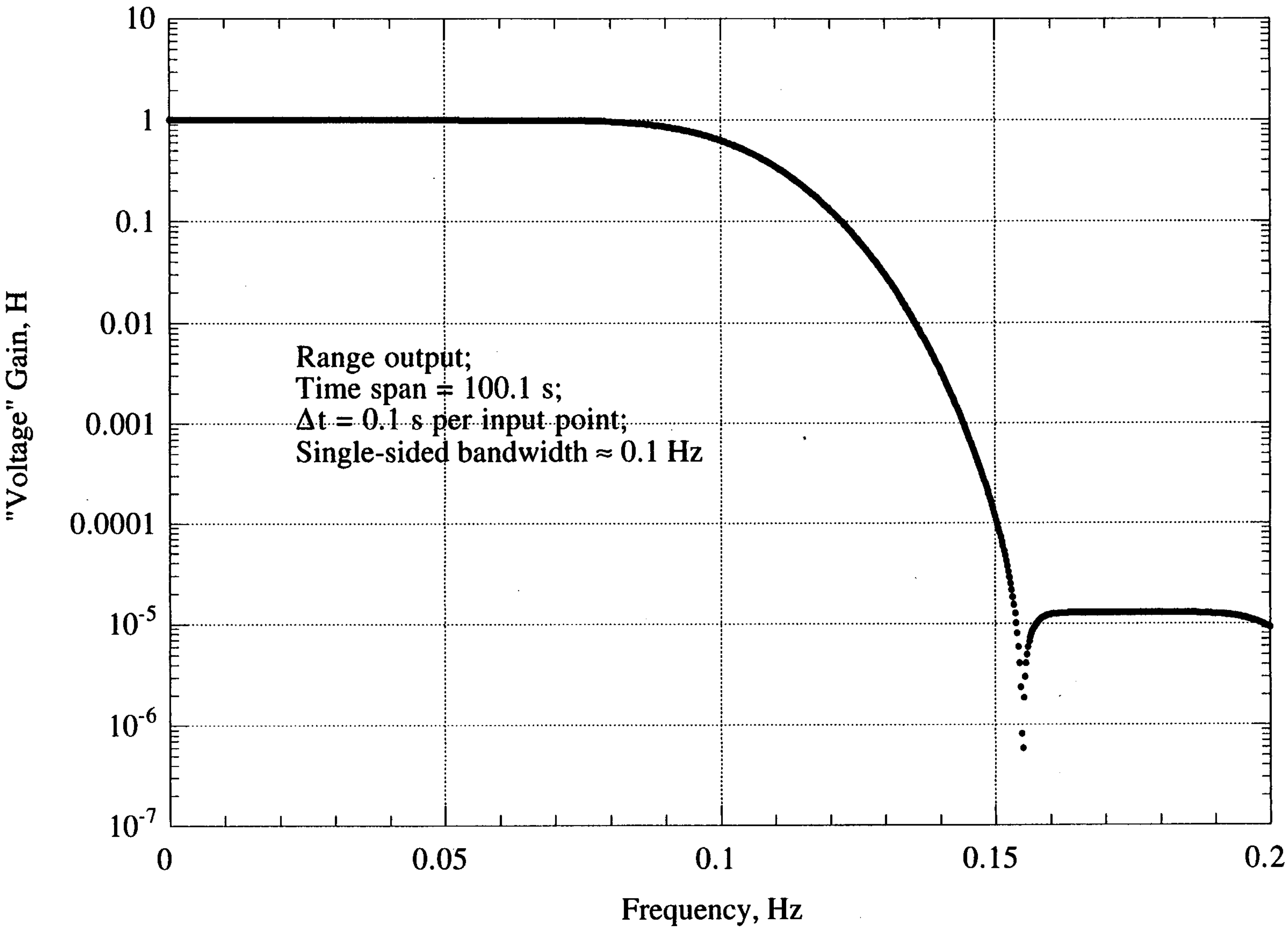

Figure A-10 An Example of a CR7 Digital Filter in the Frequency Domain

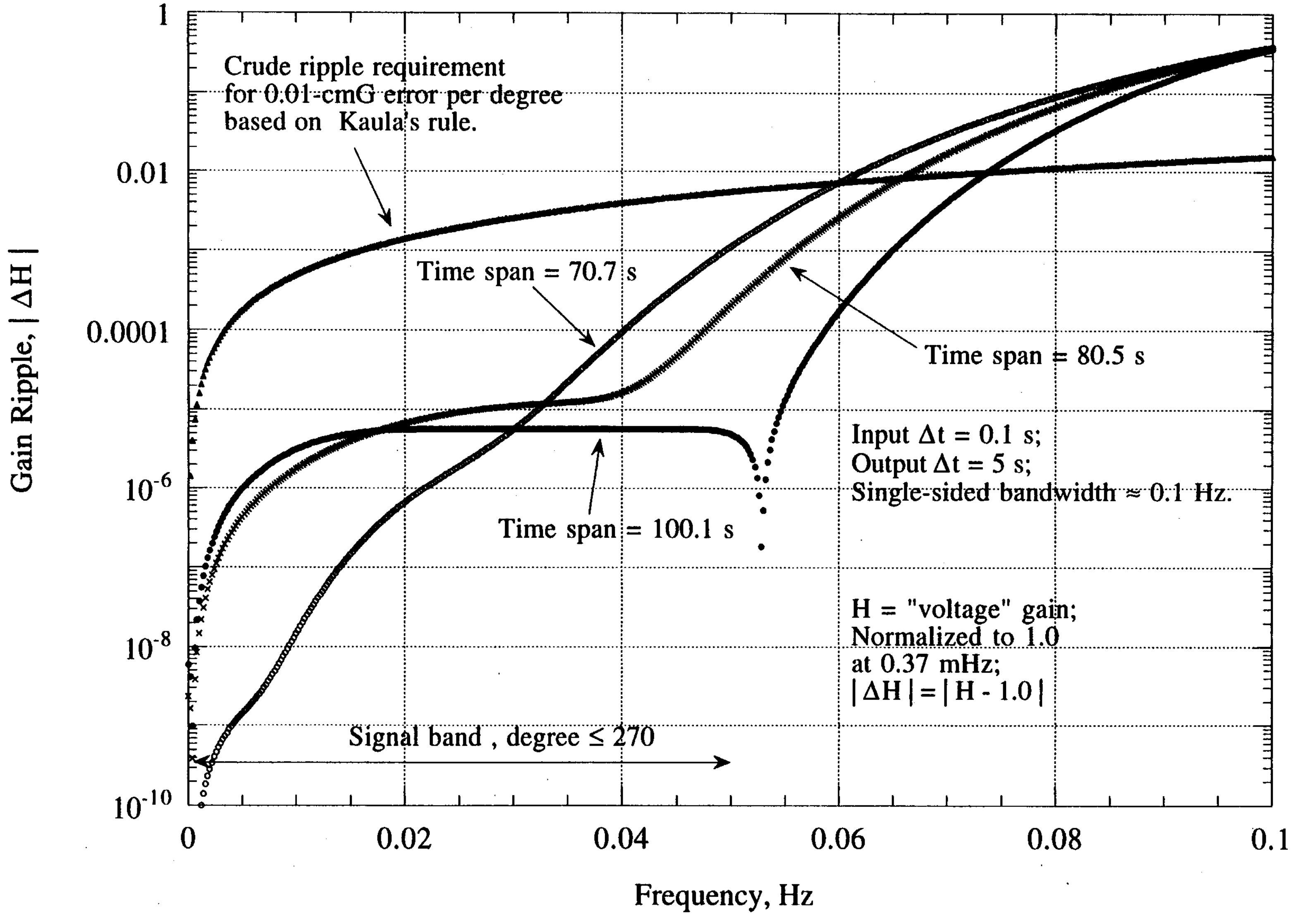

Figure A-11 Gain Ripple for Selected CR7 Digital Filters

A.3.7 Examples of Digital Filters for the Range-Rate and Acceleration Observables

The unnormalized range-rate and acceleration filters for use in Eqs. (A.11) and (A.12) are generated with the following DFTs:

$$\dot{F}_n = \sum_{k=-N_h}^{N_h} -(2\pi f_k) H_k \sin(2\pi k n / N_f) \qquad \text{for } |n| \le N_h \tag{A.20}$$

for the range-rate filter and

$$\ddot{F}_n = \sum_{k=-N_h}^{N_h} -(2\pi f_k)^2 H_k \cos(2\pi k n / N_f) \qquad \text{for } |n| \le N_h \tag{A.21}$$

for the acceleration filter. In these expressions, the frequency is given by $f_k = k / T_f$ and H_k is given by Eq. (A.17). When these filters are generated with Eqs. (A.20) and (A.21) on the basis of the H_k values used above for the range filter and normalized at the J2 frequency in a manner similar to that used for F_n, one obtains the time-domain differentiating filters presented in Figs. A-12 and A-13. Note that these filters are essentially the time derivatives of F_n, as implied by Eqs. (A.18), (A.20) and (A.21). Figs. A-14 and A-15 present the frequency response for these range-rate and acceleration filters, respectively. As one would expect, the frequency response below 36 mHz is very nearly equal to $2\pi f$ for the range-rate filter (when applied to antisymmetric sine input) and $-(2\pi f)^2$ for the acceleration filter (when applied to symmetric cosine input), with both filters applying a frequency cutoff at approximately 0.1 Hz, as required. The fractional gain ripple for each of these filters is essentially the ripple shown in Fig. A-11.

A.3.8 Examples of Other CRN Digital Filters

To illustrate the wide variety of options provided by CRN filters, Fig. A-11 plots the gain ripple generated when the CR7 filter of that figure is assigned other (shorter) filter time spans, while Fig. A-16 presents the gain ripple generated by three CR5 filters. In Fig. A-11, the ripple error increases at the upper edge of the nominal "allowed" signal band (i.e., up to 0.05 Hz) as filter time span decreases. To minimize damage caused by bad input points, the shortest possible time span that is compatible with other requirements should be selected. In Fig. A-11, the time span of 70.5 s produces excellent gain ripple (on the order of 1 μmG or less) up to approximately 0.036 Hz (degree 200) but ripple becomes relatively large ($\approx$ 10 μmG or less) near 0.05 Hz (degree 270).

The CR5 filters illustrated in Fig. A-16 would allow a somewhat shorter time span than the CR7 filters if one opted for considerably less margin with respect to gain ripple. For example, a time span of 50.5 s for the indicated CR5 filter would provide gain ripple on the order of 10 μmG or less up to 0.036 Hz (degree 200) and on the order of 30 μmG or less up to 0.05 Hz (degree 270).

Selection of the best filter parameters will depend strongly on the prevalence of bad input points, which are expected to be infrequent but may not be. Since these filtering operations are not real time, are on the ground, and can be easily changed, filter characteristics can be optimized later on the basis of the quality of actual data. However, the preceding examples indicate that it should be possible to implement a digital filter that has satisfactory-to-excellent ripple performance and a reasonably short time span (e.g., on the order of 50 to 100 s for 5 s output points).

As noted above, ripple performance for other filters can be approximately obtained by simply scaling the results in Figs. A-11 and A-16. For example, a CR7 filter corresponding to the 70.7-s filter in Fig. A-11, but with 10-s output points rather than 5-s, would possess a time span of 140.7 s and a bandwidth of approximately 0.05 Hz. This would lead to approximately the ripple shown in Fig. A-11 for the 70.7-s case but with the frequency axis divided by two (i.e., 0.1 Hz in Fig. A-11 becomes 0.05 Hz).

A.3.9 An Example of Processing Random Noise with the Acceleration Filter

A double-differentiating digital filter has been applied to (white) Gaussian random noise, uncorrelated from point to point, with a 1-σ magnitude of 3.16 μm on the 10-S/s points. Noise of this magnitude is equivalent to a ("single-sideband") spectral noise density of 1 μm^2 Hz^{-1} on the input range values. A sequence of 2×10^5 5-s output points spanning approximately 13 days was generated on the basis of approximately 10^7 0.1-s input points. In this example, the filter bandwidth is set to 0.05 Hz, which would be appropriate for a 0.1 Hz output rate. However, the output rate is changed from 0.1 Hz to 0.2 Hz in order to reveal the 0.05-Hz cutoff of the filter. A DFT was then applied to this output sequence over 13 days to obtain the spectral density. The resulting spectrum, presented in Fig. A-17, shows the expected $(2\pi f)^2$ dependence for frequencies between approximately 2 and 18 mHz. Below 2 mHz, a noise floor appears at the level of approximately 10^{-4} $\mu m/(s^2 \sqrt{Hz})$ as a result of the sidelobes of the DFT sinx/x filter reaching out and collecting power near 0.05 Hz. This noise floor, which decreases as the square root of DFT (or multiparameter fit) integration time, is at a sufficiently low level that gravity-recovery performance is not significantly degraded.

112

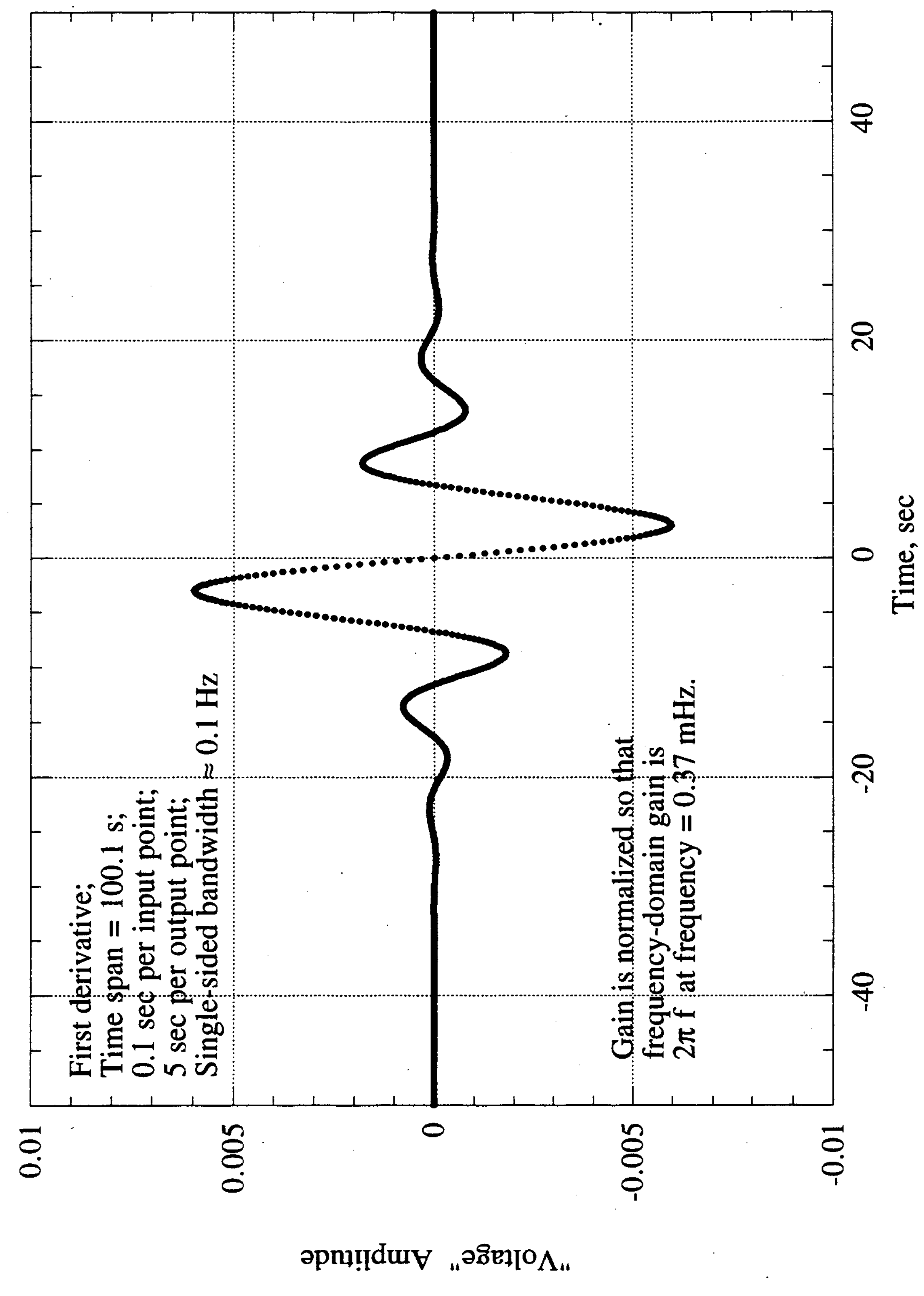

Figure A-12 Example of a CR7 Differentiating Digital Filter in the Time Domain

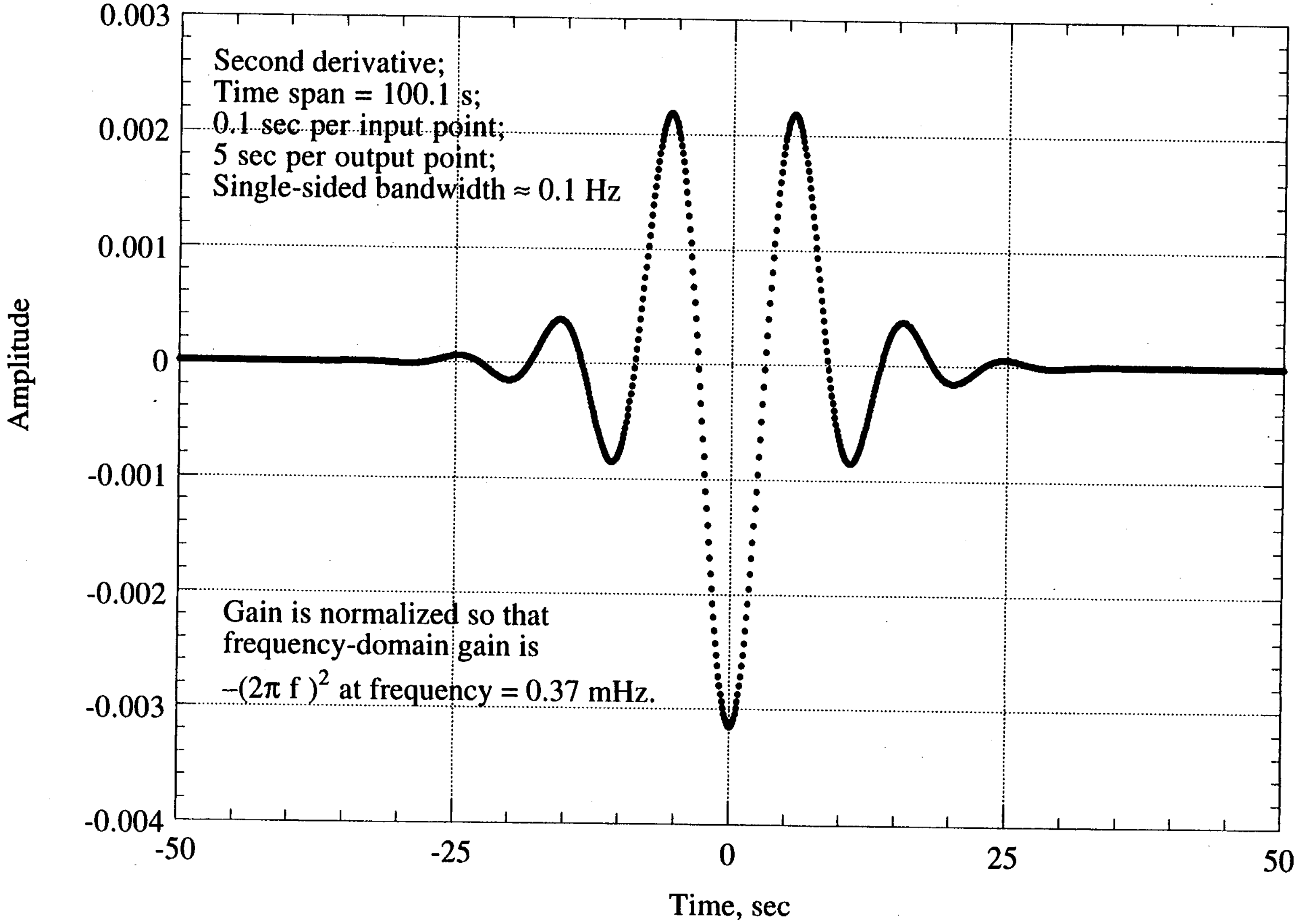

Figure A-13 Example of a CR7 Double Differentiating Digital Filter in the Time Domain

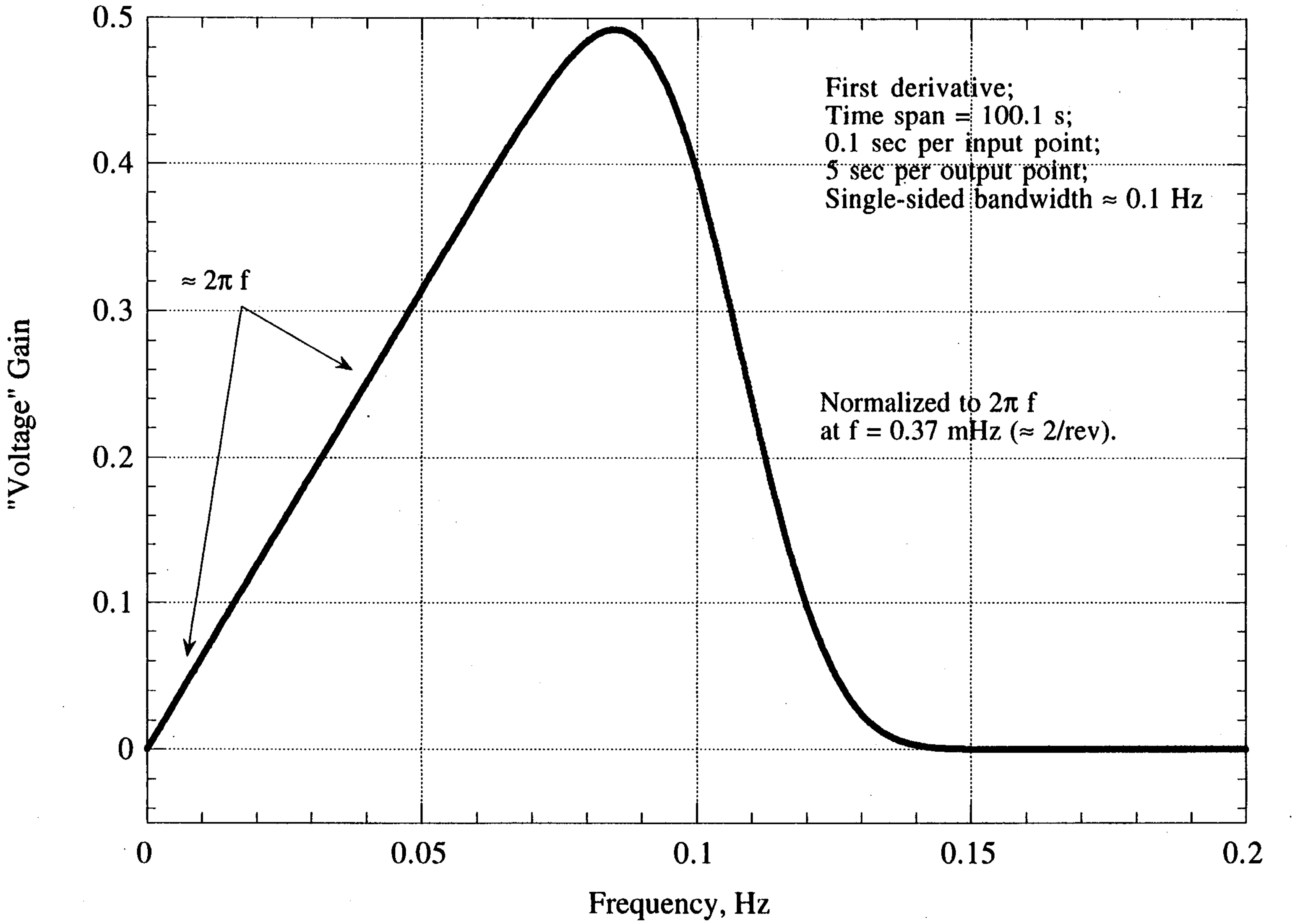

Figure A-14 An Example of a CR7 Differentiating Digital Filter in the Frequency Domain

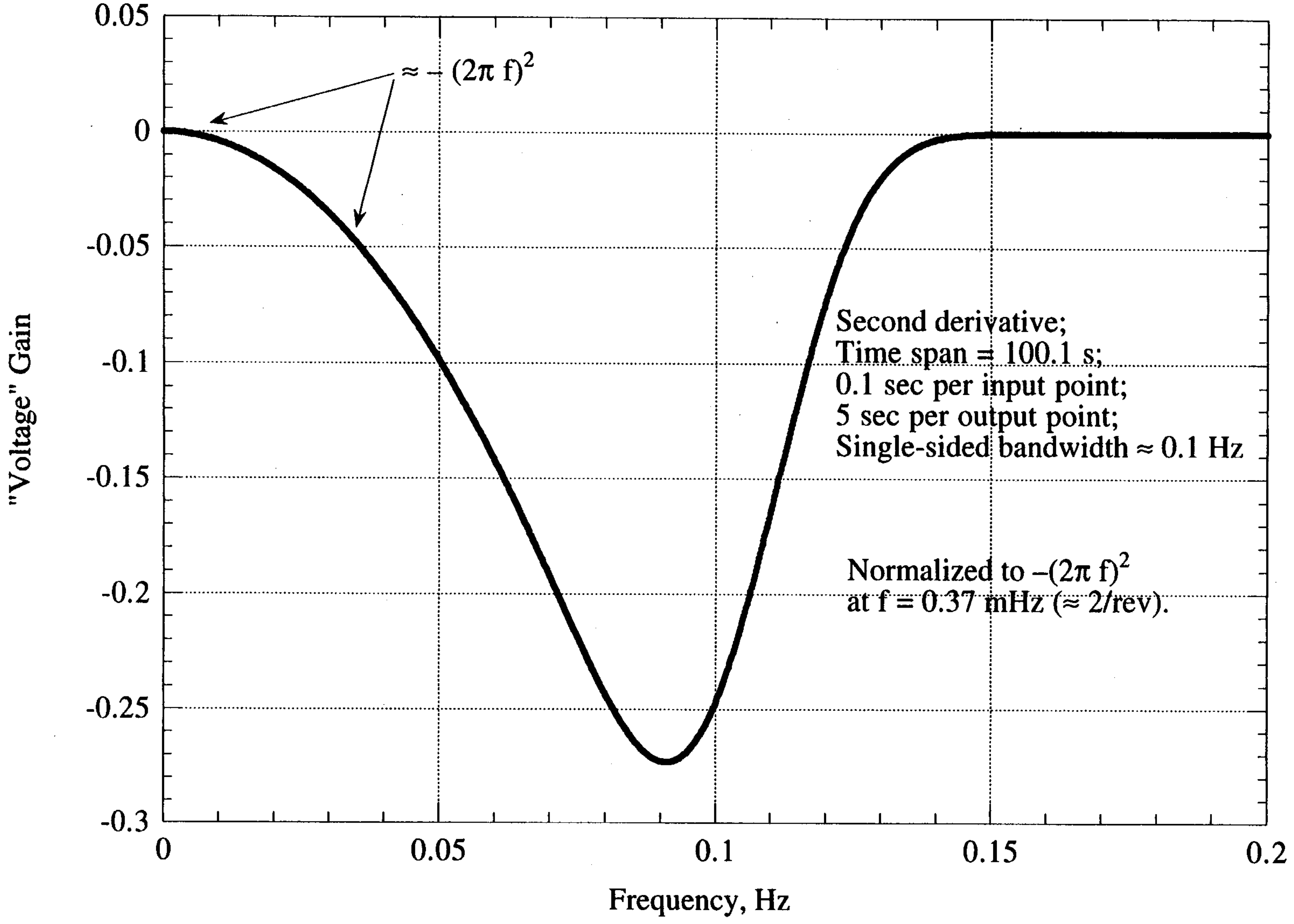

Figure A-15 An Example of a CR7 Double-Differentiating Digital Filter in the Frequency Domain

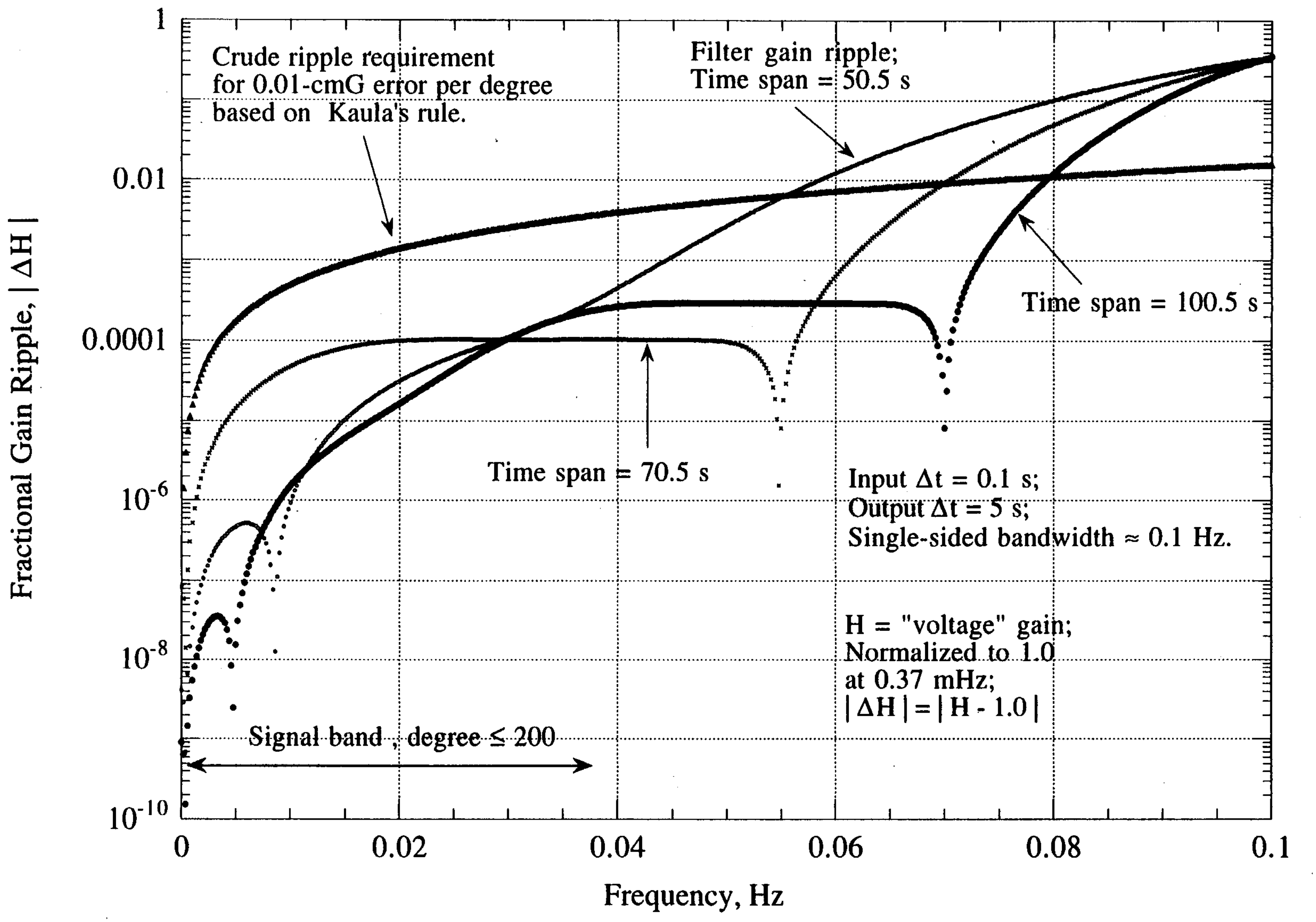

Figure A-16 Fractional Gain Ripple for Three CR5 Digital Filters

118

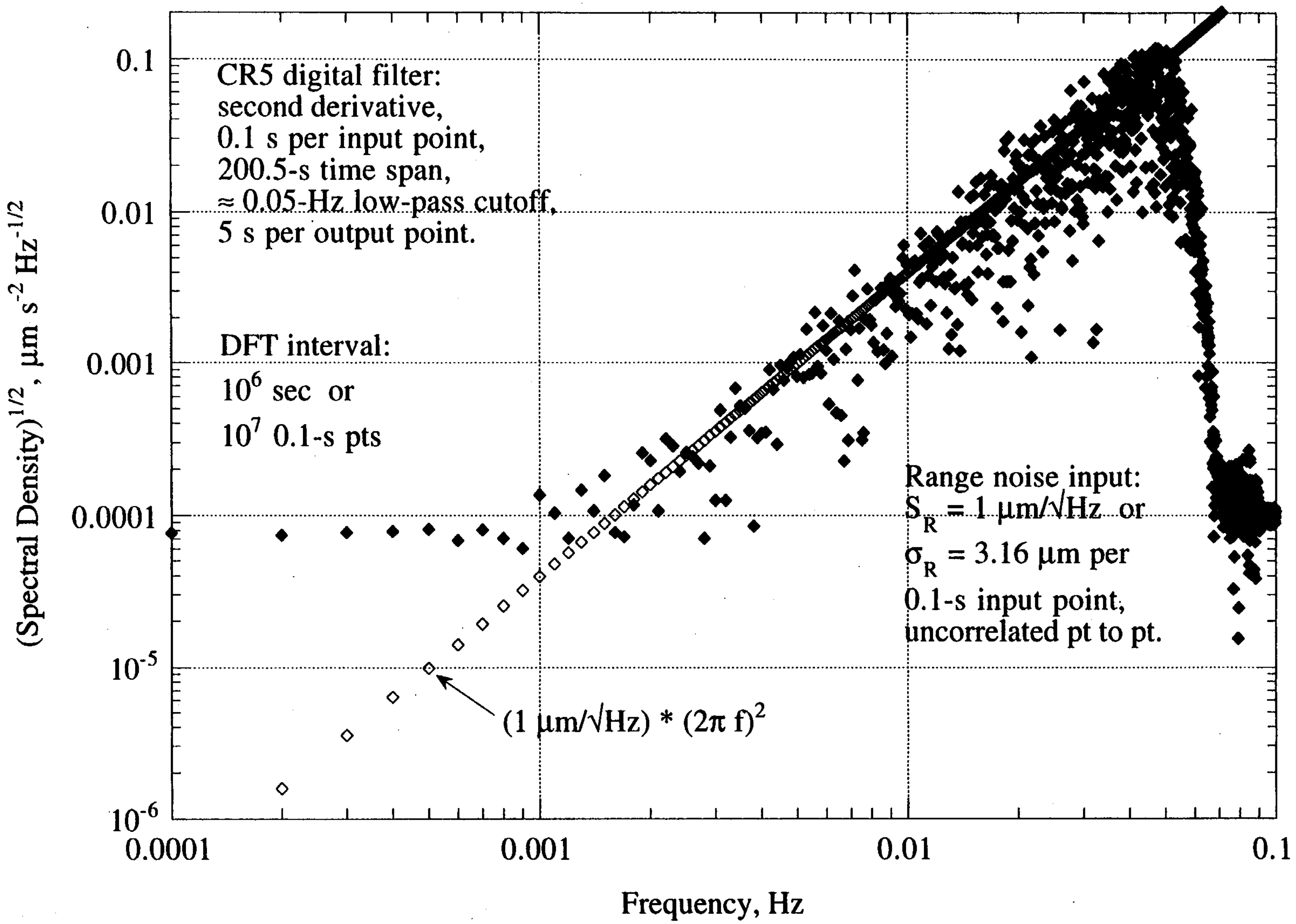

Figure A-17 An Example of Spectral Density for Filtered, Double-Differentiated White Random Noise

APPENDIX B

RANDOM NOISE SPECTRA FOR THREE NOISE SOURCES

B.1 System Noise

Measurements of K-band phase will be corrupted by system noise (background plus thermal noise). System noise, which is white and reasonably predictable, is often specified in terms of the ratio C/N_0 in units of dB-Hz, where C is received signal power and N_0 is the spectral noise density at the receiving satellite. C/N_0 can be thought of as power SNR in the form of input signal power divided by noise power in a 1 Hz bandwidth. For illustration, it is assumed in this report that a C/N_0 of approximately 69 dB-Hz can be expected for an intersatellite range of 200 km.

When processing losses (≈ 3 dB) are taken into account, the voltage SNR expected for a BlackJack receiver operating in the quadrature down-conversion and sampling mode is given in terms of C/N_0 by

$$\mathrm{SNR_V} \approx \left[\frac{C}{N_0} T_{obs}\right]^{1/2} \tag{B.1}$$

where T_{obs} is the integration time for the observation. For $C/N_0 = 69$ dB-Hz and an observation time of 1 s, $\mathrm{SNR_V}$ at each satellite is 2800. The 1-σ phase error is given in terms of $\mathrm{SNR_V}$ by

$$\sigma_\phi = \frac{1}{2\pi \mathrm{SNR_V}} \tag{B.2}$$

in units of cycles. For $\mathrm{SNR_V} = 2800$, the 1-s system noise error in phase is approximately 57 μcyc, which is equivalent to 0.53 μm for a wavelength of 0.94 cm at 32 GHz.

System noise is uncorrelated between satellites and between RF bands. Thus, when the phase values measured by the two satellites are combined to generate dual-1-way range, as prescribed by Eq. (3.14), the overall system noise is reduced by a factor of √2. (See Eq. (3.25) where √2 arises in the numerator and a factor of two in the denominator). Further, in the ionosphere-correction operation given by Eq. (3.30), the dual-band combination of the two ranges measured at K-band and Ka-band increases system-noise error in range by a factor equal to

$$M_I = \left[\left(\frac{f_K^2}{f_K^2 - f_{Ka}^2}\right)^2 + \left(\frac{f_{Ka}^2}{f_K^2 - f_{Ka}^2}\right)^2\right]^{1/2} \tag{B.3}$$

if one assumes the SNRs are the same for the two bands. For the frequencies of 24 and 32 GHz, this ionosphere-calibration noise-amplification factor is $M_I = 2.62$. (A better use of transmit power would be to allot less power to the lower frequency and more to the higher frequency. This option is not considered here.) Thus, the final system-noise error in range for a 1-s integration time is

$$\sigma_R \approx \left(\frac{2.62}{\sqrt{2}}\right) * 0.53\ \mu\mathrm{m} \approx 1\ \mu\mathrm{m} \tag{B.4}$$

Since system noise is white, this 1-σ error for a 1-s integration time corresponds to a flat spectral noise density (SND) of 1 $\mu m^2/Hz$ over the range-signal band. This is a "single-sideband" value, signifying no "fold-over" of the negative-frequency part of the spectrum has been applied.

When biased range measurements are passed through the differentiating filters (see Appendix A) to obtain range rate and range acceleration, the resulting noise spectra are modified by factors of $2\pi f$ and $(2\pi f)^2$, respectively, where f is the frequency of the relevant spectral component. Thus, the square roots of the SNDs are approximately given by

$$\sqrt{S_R(f)} = 1\ \mu m/\sqrt{Hz} \quad \text{for biased range} \tag{B.5}$$

$$\sqrt{S_{\dot{R}}(f)} = (2\pi f) * 1\ \mu m/\sqrt{Hz} \quad \text{for range rate} \tag{B.6}$$

$$\sqrt{S_{\ddot{R}}(f)} = (2\pi f)^2 * 1\ \mu m/\sqrt{Hz} \quad \text{for range acceleration} \tag{B.7}$$

These expressions have currently been adopted as the nominal values for single-sideband SND for the system-noise error in the three primary observables. (As explained in Appendix A but not accounted for in Eq. (B-7), the system-noise error for the acceleration observable can be "degraded" in fits to estimate gravity, by an effective fit-interval-specific "noise floor.") A plot of the system-noise √SND for acceleration, as given by Eq. (B.7), is presented in Fig. B-1 along with the √SNDs for the USOs and accelerometers discussed below. Differentiation applied to obtain range rate and acceleration changes the white-noise spectrum for range to "blue-noise" spectra, with the acceleration spectrum particularly biased toward higher frequencies. Figs. B-2 and B-3 present corresponding √SNDs for the range-rate and range observables, respectively.

B.2 Ultrastable Oscillator

As indicated in Eqs. (3.1) and (3.2), oscillator instabilities corrupt K-band phase measured at each satellite. Oscillator phase stability over long time scales is typically specified in terms of Allan deviation, $\sigma_y(\tau)$, where τ is the integration time. For illustration, it is assumed in this report that Allan deviation for the GRACE ultrastable oscillators (USOs) is given by $\sigma_y(\tau) = 2 \times 10^{-13}$, 1×10^{-13}, 1×10^{-13}, and 2×10^{-13} for integration times of $\tau = 1$, 10, 100, and 1000 s, respectively. For the error analysis of this report, it is advantageous to express phase instability in the frequency domain. A number of methods can be used to map the Allan-deviation specifications in the time domain to the frequency domain to obtain the corresponding spectral noise density for phase. The following method avoids the discontinuities and artificial boundaries in integration time found in some approaches.

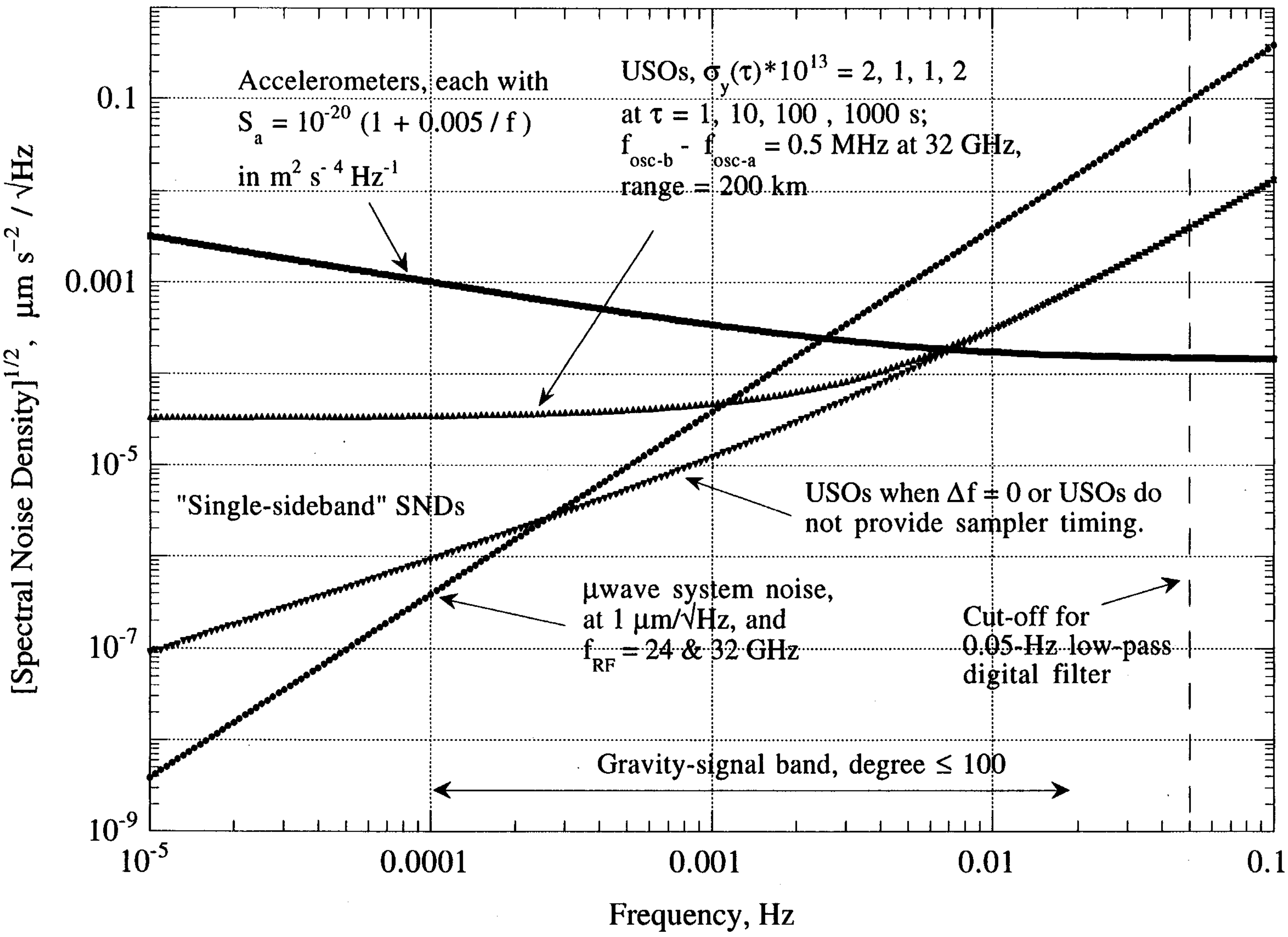

Figure B-1 Spectral Noise Densities for Three Noise Sources for the Acceleration Observable

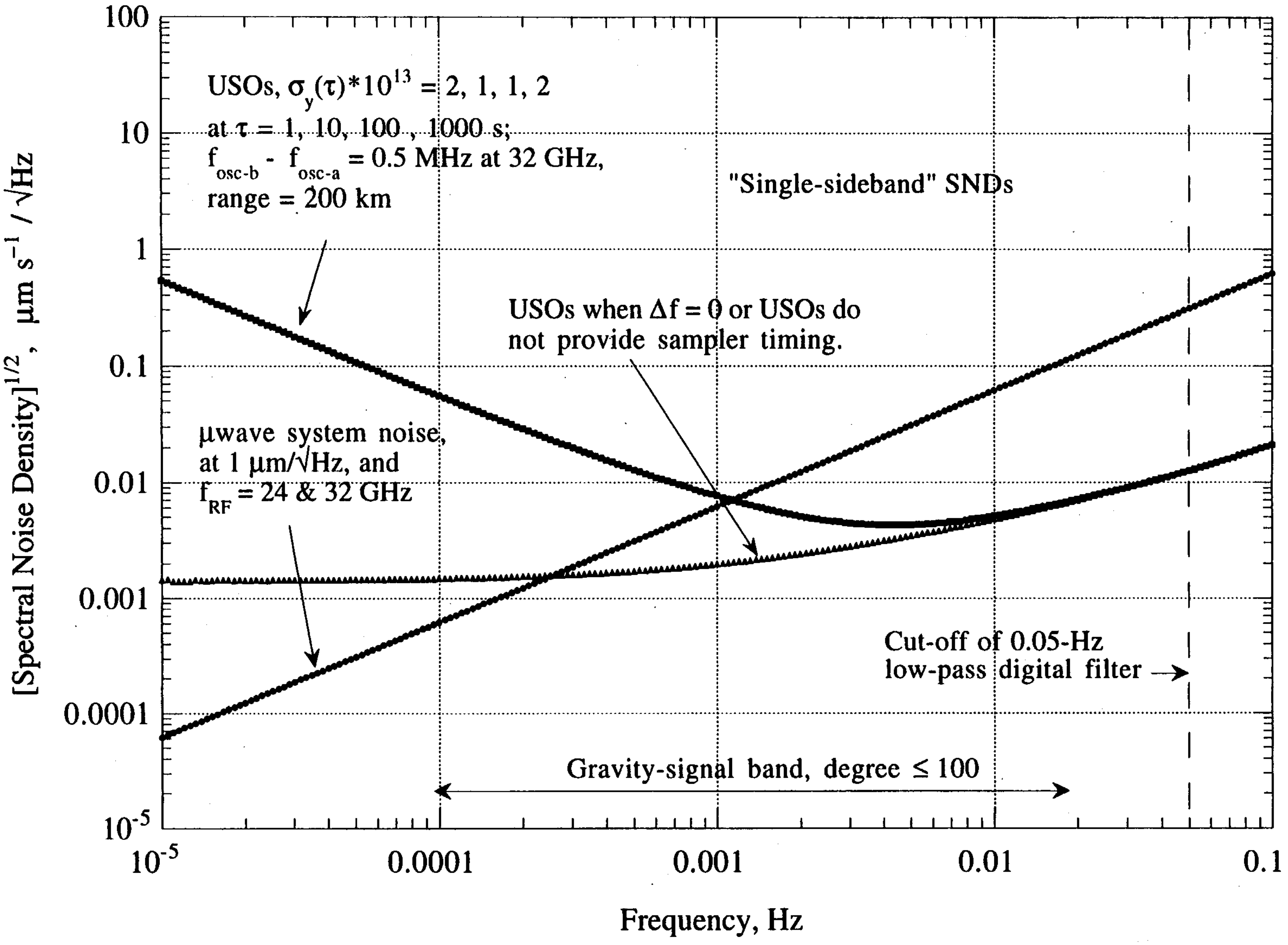

Figure B-2 Spectral Noise Densities for Two Noise Sources for the Range-Rate Observable

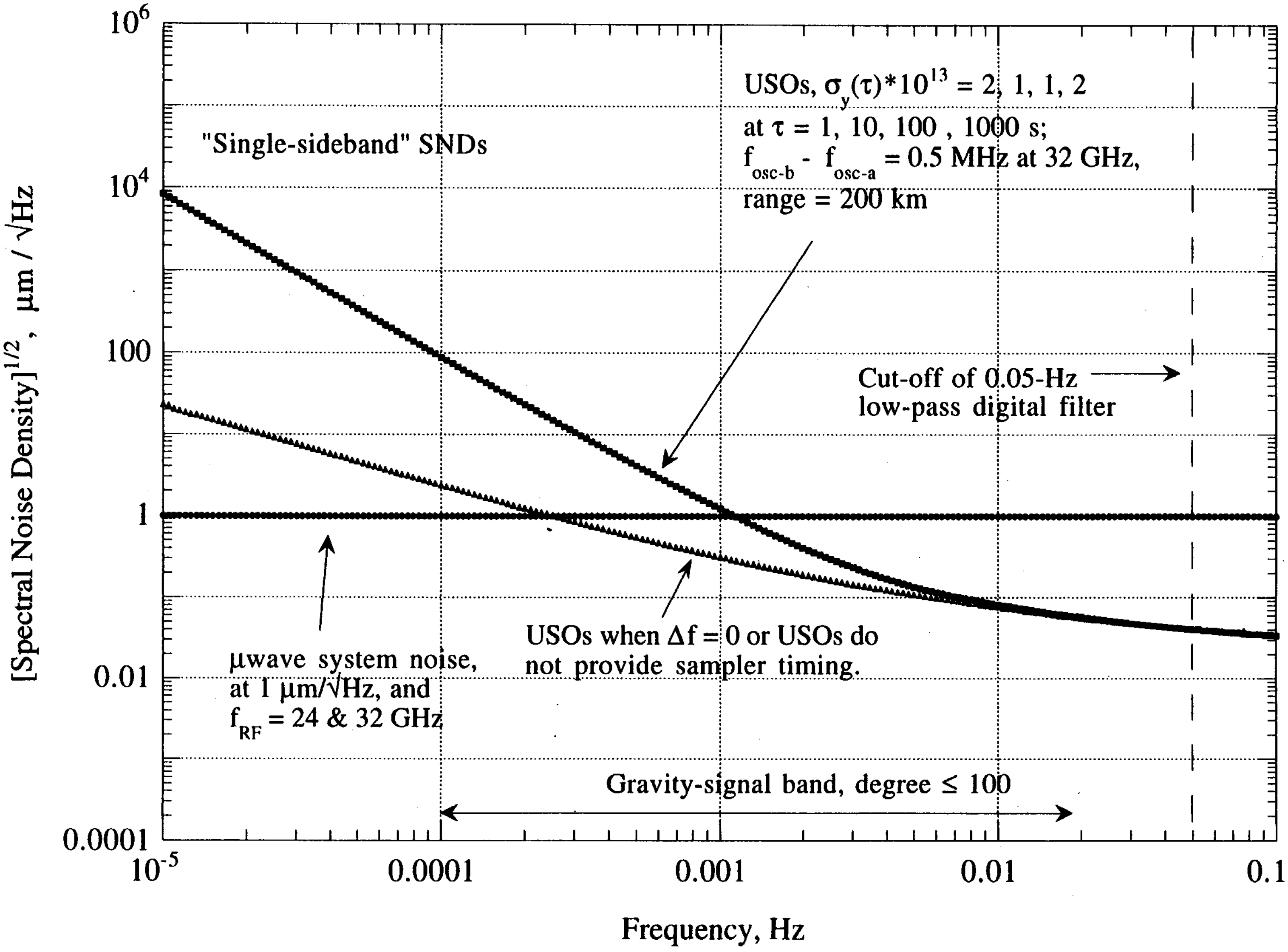

Figure B-3 Spectral Noise Densities for Two Noise Sources for the Range Observable

Oscillator instability is often expressed in terms of single-sideband (SSB) phase noise, which is the power spectrum of random phase deviations (η_p in Eq. (3.1)) relative to "ideal" linear time behavior. "Single sideband" denotes specification of the power spectrum for positive frequencies, without the doubling sometimes applied to arrive at "double-sideband" phase noise that accounts in advance for expected "negative-frequency fold-over." SSB phase noise at frequency f can be approximated as an f^{-n} expansion (e.g., see [7]) of the form:

$$S_\phi(f) = h_0 + \frac{h_{-1}}{f^1} + \frac{h_{-2}}{f^2} + \frac{h_{-3}}{f^3} + \frac{h_{-4}}{f^4} \qquad \text{(B.8)}$$

where S_ϕ is in units of rad^2/Hz and pertains to a specified oscillator output frequency (e.g., 5 MHz). The oscillator phase noise at frequency f in bandwidth Δf can be computed as $\sqrt{S_\phi(f)\,\Delta f}$ in radians.

If USO stability is specified in terms of Allan deviation, a method is needed to derive the h_{-n} coefficients from the given Allan deviation. Allan variance can be expressed in terms of h_{-n} coefficients as

$$\sigma_y^2(\tau) = \frac{1}{f_0^2}\left[\frac{h_0 f_h}{2.565^2}\tau^{-2} + \frac{h_{-1}[2.184 + \ln(f_h\tau)]^2}{2.565^2}\tau^{-2} + h_{-2}\tau^{-1} + 1.665^2 h_{-3} + 3.63^2 h_{-4}\tau\right] \qquad \text{(B.9)}$$

where f_0 is the oscillator output frequency and f_h (typically $\approx$ 1 Hz) is the low-pass bandwidth of the measurement system. In the calculation below, f_0 = 5 MHz, a typical USO value, will be adopted as the oscillator frequency. (The specific value adopted for the reference frequency in the following computation does not effect the final value for oscillator stability at RF since the scaling to RF always leads to the same result.) Eq. (B.9) has been constructed from the entries in Table 2-2 on p. 74 in reference [7], under the assumption that the f^{-n} components are uncorrelated.

Note in Eqs. (B.8) and (B.9) that the h_0 term dominates at high frequencies (e.g., $f > 1$ kHz) and small integration time (e.g., $\tau < 0.001$ s). When the smallest specified integration time is 1 s, Allan deviation values supply little information concerning h_0. Fortunately, the h_0 coefficient can be relatively easily separated out by directly measuring the spectral noise density (the "noise floor") at a "high" frequency (e.g., at f = 100 kHz). A typical value for h_0 is approximately –155 rad^2/Hz for f_0 = 5 MHz for GRACE-type USOs. In this analysis, the 1/f term in Eq. (B.9) can be neglected since this term is small for quartz USOs [R. L. Sydnor, private communication, 1996]. If the 1/f term is neglected and h_0 is known, Eq. (B.9) has three unknowns, h_{-2}, h_{-3}, and h_{-4}. Since the Allan deviation has been specified for four integration times, one can solve for the three unknowns.

For the Allan deviation values given at the beginning of this subsection, this approach results in the following expression for the USO SSB phase noise in units of rad^2/Hz:

$$S_\phi = 3.16\text{x}10^{-16} + \frac{8.38\text{x}10^{-13}}{f^2} + \frac{5.74\text{x}10^{-14}}{f^3} + \frac{6.39\text{x}10^{-17}}{f^4} \qquad \text{(B.10)}$$

where the implicit reference frequency is the adopted oscillator frequency of 5 MHz and f in Hz. For the Ka band, the SSB phase noise in Eq. (B.10) is translated from 5 MHz up to 32 GHz by multiplying by (32 GHz/5 MHz)^2 . A plot of the SSB phase noise in Eq. (B.10) is shown in Fig. B-4, for both 5 MHz and 32 GHz reference frequencies.

When dual-1-way range is computed by combining the two 1-way K-band phases extracted by the two satellites, as prescribed in Eq. (3.14), a massive improvement in oscillator stability is obtained. The dual-1-way combination results in the effective time-domain processing of oscillator instability given by Eq. (3.36), which is equivalent to the frequency-domain filter given by Eq. (3.38). Since oscillator noise is nondispersive (i.e., has the same effect in microns for the 24 and 32 GHz bands), no adjustment is necessary to account for the dual-band combination that corrects for ionosphere. Thus, the SND of total effective oscillator noise for range is computed by multiplying Eq. (B.10) by the square of Eq. (3.38), dividing by $(2\pi)^2$ to convert from rad^2 to cyc^2, multiplying by (32 GHz/5 MHz)^2 to translate to K-band, and multiplying by two to account for two oscillators:

$$\boxed{S_R(f) = \frac{1}{2}\left|\left(\frac{f_q}{f_p} - e^{-2\pi i f\tau}\right)\right|^2 \left[\, 0.029 + \frac{77}{f^2} + \frac{5.3}{f^3} + \frac{0.0059}{f^4} \right]} \tag{B.11}$$

where the units are μm^2/Hz, f is in Hz, and f_p, f_q are the two tone frequencies for the band being considered (at K band or at Ka band). It is expected that f_p and f_q will have a separation of approximately 0.5 MHz. The wavelength λ_e in Eq. (3.38) is omitted in Eq. (B.11) since wavelength is implicitly accounted for in the coefficients.

The corresponding SNDs for range rate and acceleration are computed as they are for the system-noise case, namely by multiplying the total effective SND for range in Eq. (B.11) by $(2\pi f)^2$ for range rate and by $(2\pi f)^4$ for range acceleration. The resulting USO SNDs are plotted in Figs B-1, B-2, and B-3 in the form of $\sqrt{\text{SND}}$, given a nominal tone separation at RF of $\Delta f \equiv f_q - f_p = 0.5$ MHz. The strong increase in the USO SND for range at low frequencies is avoided in the acceleration observable as a result of differentiation (i.e., $\sqrt{\text{SND}} \rightarrow$ constant for acceleration vs. $\sqrt{\text{SND}} \rightarrow f^{-2}$ for range).

For reference, Figs. B-1, B-2, and B-3 also plot USO SND for the case when sampler timing is not supplied by the USOs but is derived from concurrent GPS measurements of pseudorange and phase. This is mathematically accomplished by setting $\tilde{s}_p - \tilde{s}_q = 0$ in Eq. (3.19) or $f_p = f_q$ in Eq. (B.11). This approach removes USO noise from timetags but introduces a new error source, the noise caused by the aforementioned GPS-derived timetags (i.e., in $\tilde{s}_p - \tilde{s}_q$) interacting with $f_q - f_p$. This error source is not analyzed in this report. Note that USO SND at low frequencies is proportional to $1/f^2$ for range and not $1/f^4$, when time tag stability is derived from GPS measurements.

The preceding analysis suggests that it would be advantageous, at least in some applications, to specify oscillator stability completely in terms of SSB phase noise rather than in terms of both SSB phase noise (for "high frequencies") and Allan deviation (for "low frequencies"), as is often the case. Provided the expansion in Eq. (B.8) is a reasonably accurate representation of SSB phase noise over the relevant frequency range, only one expression with a small number of coefficients is required to describe both short term and long term behavior over a wide range. Such a specification would eliminate the need for carrying out the transformation in this subsection, when frequency-domain specification and analysis are the goal.

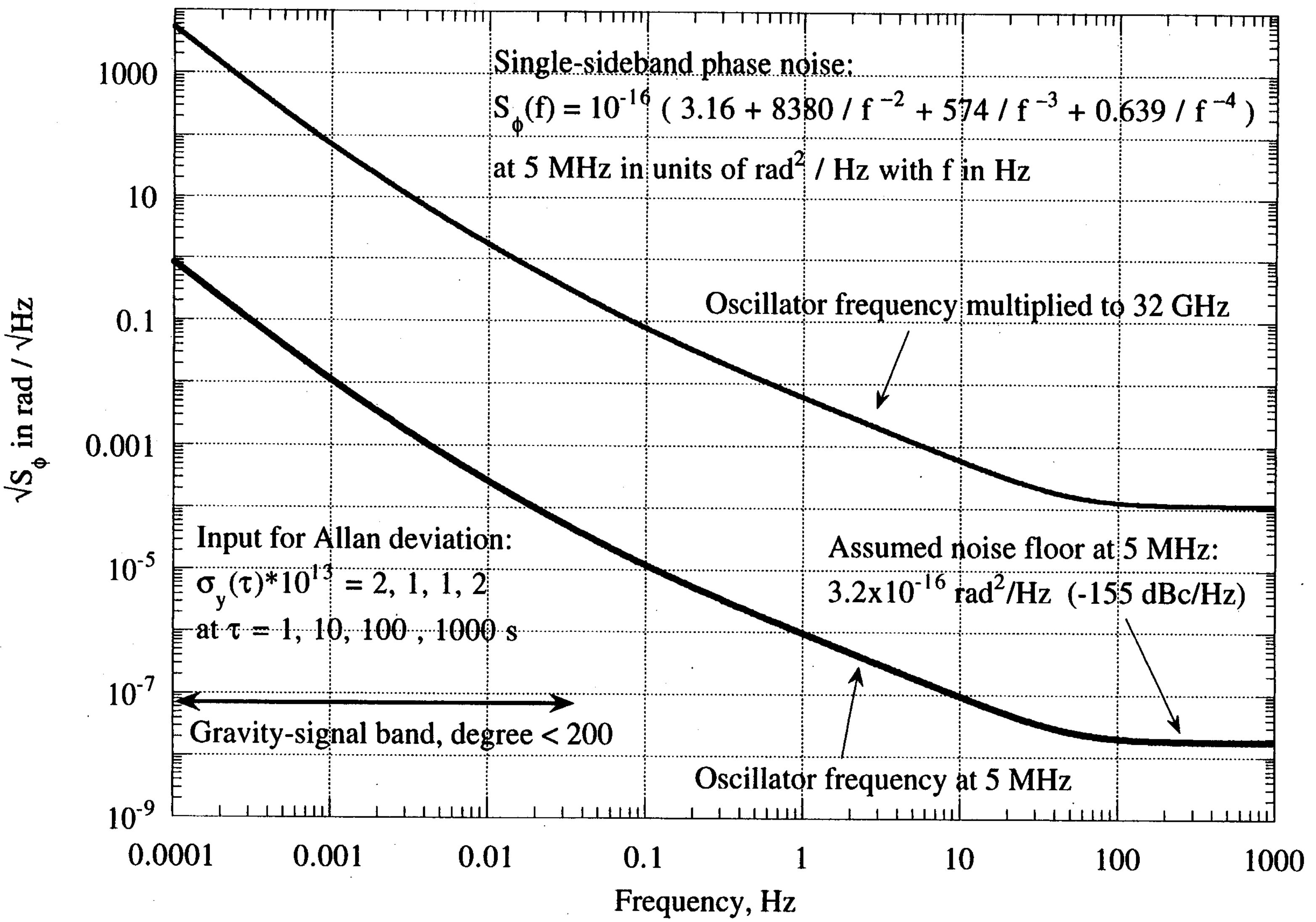

Figure B-4 Estimate of Single-Sideband Phase Noise for GRACE Ultrastable Oscillator

B.3 Accelerometer

The random noise on acceleration values measured by a given accelerometer along either of its two most accurate axes is specified to have a SND given by

$$S_a(f) = 10^{-20}(1 + 0.005/f) \qquad (B.12)$$

in units of $m^2\,s^{-4}Hz^{-1}$ with f in units of Hz. Current plans call for one of the sensitive axes to be aligned along the line-of-sight (LOS) between GRACE satellites and the other in the radial direction. The third axis, which is 5 to 15 times noisier in the gravity-signal band in terms of √SND, is specified to have a SND of

$$S_a(f) = 10^{-18}(1 + 0.1/f) \qquad (B.13)$$

in units of $m^2\,s^{-4}Hz^{-1}$ with f again in Hz. This axis will be aligned cross track (i.e., approximately in the longitude direction).

The LOS "drag" component, $\ddot{\mathbf{R}}_d \cdot \hat{\mathbf{R}}$, is obtained by taking the difference the two accelerometer outputs, one for each satellite, measured along the line-of-sight direction. Thus, provided proper sampling and filtering of the accelerometer measurements is carried out, the accelerometer noise from each satellite enters $\ddot{\mathbf{R}}_d \cdot \hat{\mathbf{R}}$ directly as given by Eq. (B.12). Since noise from two accelerometers corrupts $\ddot{\mathbf{R}}_d \cdot \hat{\mathbf{R}}$, the noise in Eq. (B.12) is multiplied by two to obtain the total accelerometer noise. Thus, the total accelerometer noise in $\ddot{\mathbf{R}}_d \cdot \hat{\mathbf{R}}$ is given by

$$S_a(f) = 2x10^{-20}(1 + 0.005/f) \qquad (B.14)$$

in units of $m^2\,s^{-4}Hz^{-1}$ with f in units of Hz. Fig. B-1 plots √SND for $\ddot{\mathbf{R}}_d \cdot \hat{\mathbf{R}}$ obtained from Eq. (B-14).

The accelerometer SNDs for range rate and range have not been computed for Figs. B-2 and B-3 since accelerometer output is processed differently. That is, the three-dimensional acceleration vector measured by an accelerometer is, in effect, first mapped from satellite coordinates to inertial coordinates, integrated in time (once for range rate and twice for range) and then used in the computation of range rate and range. Such mapping has not been carried out for this report since the focus here is the acceleration observable.

B.4 Combined Noise

For acceleration as shown in Fig. B-1, system noise is dominant at frequencies greater than approximately 2.5 mHz , while accelerometer noise is dominant for frequencies less than 2.5 mHz. The total random noise spectrum, computed as the root-sum-square (RSS) of the plotted quantities, deviates substantially from a white (flat) spectrum, with the RSS varying by as much as a factor of 30 ($3x10^{-4}$ to 10^{-2} μm/s^2) over the gravity-signal band (i.e., approximately 0.1 to 18 mHz up to degree 100) indicated in the plots.

For the acceleration observable in Fig. B-1, the effective √SND that would be generated when a quadratic fit (see the modulo'ed effective filter in Fig. A-6) operates on 1-μm $Hz^{-1/2}$

system noise is approximately equal to the constant value of $0.27\ \mu\text{m}\,\text{s}^{-2}\text{Hz}^{-1/2}$, spanning –0.05 to +0.05 Hz (assuming 10-s output points). The increase due to modulo'ed USO noise would be minimal since out-of-band USO noise is much smaller than out-of-band system noise in this example. Thus, depending on the frequency point in the signal band, quadratic-fit processing would increase the effective in-band $\sqrt{\text{SND}}$ for acceleration on the order of 10- to 1000-fold relative to in-band system noise and USO noise in Fig. B-1.

A similar analysis can be carried out for range rate in Fig. B-2. The effective $\sqrt{\text{SND}}$ that would be generated when a quadratic fit (see the modulo'ed effective filter in Fig. A-5) operates on $1\text{-}\mu\text{m}\,\text{Hz}^{-1/2}$ system noise is approximately equal to the constant value of $0.35\ \mu\text{m}\,\text{s}^{-1}\text{Hz}^{-1/2}$, again spanning –0.05 to +0.05 Hz (assuming 10 s output points). The increase due to modulo'ed USO noise would again be minimal in this example. Thus, depending on the frequency point in the signal band (0.18 to 18 mHz), quadratic-fit processing would increase the effective in-band $\sqrt{\text{SND}}$ for range rate on the order of 3- to 35-fold relative to in-band system noise and USO noise in Fig. B-2. In the lower 1/rev to 30/rev range ($\approx$ 0.18 to 5.5 mHz) that is important for the weaker higher-order coefficients, the improvement is 10- to 35-fold.

When a corresponding analysis is carried out for the range observable, the improvement provided by digital filtering relative to a quadratic fit is less dramatic. The constant $1\text{-}\mu\text{m}\,\text{Hz}^{-1/2}$ system noise in Fig. B-3 would be increased by a factor of 1.5 (see modulo'ed effective filter in Fig. A-4), which causes total noise in the figure to increase in the upper part of the signal band by about a factor of 1.5.

To crudely assess the significance of these plots in gravity estimation, the spectral densities can be multiplied by the bandwidth effectively applied by the multiparameter, multiday fits to be carried out to obtain gravity coefficients. For a 13-day fit, the estimation process for a given gravity coefficient effectively applies to the observables a filter that can be approximately viewed as a combination of tones, each approximately 1-µHz wide. Thus, values obtained from these noise plots are to be multiplied by $\sqrt{1\ \mu\text{Hz}} = 10^{-3}\sqrt{\text{Hz}}$ to obtain at each frequency crude estimates for the effective noise that would propagate through a 13-day gravity-estimation fit. Based on Fig. B-1, the effective noise corresponding to the RSS of the SNDs for acceleration falls between $3\text{x}10^{-7}$ to $10^{-5}\ \mu\text{m/s}^2$ over the nominal signal-band range, for a 13-day fit. Similarly, the effective noise falls between 10^{-4} and $10^{-5}\ \mu\text{m/s}$ for range rate and between 10^{-3} and $3\text{x}10^{-2}\ \mu\text{m}$ for range, excluding accelerometer noise.

APPENDIX C

SIMPLE EXAMPLES OF FIT FILTERS IN THE FREQUENCY DOMAIN

To provide better understanding of multiparameter fits to extract gravity coefficients, this appendix derives the frequency-domain fit filter for a number of heuristic examples in which the amplitude of a signal is extracted in the presence of instrumental noise and/or other signals. Even though actual fits are far more complicated, these frequency-domain examples provide a number of useful insights and results concerning basic features of frequency-domain fit filters for amplitude: the selection of polynomial order and update-interval length for the instrumental model, the relation of tones in partials to mainlobes in the frequency-domain fit filter, excessive mainlobes and sidelobes, and "separation" of orbit parameters from gravity coefficients in multiparameter fits.

For GRACE measurements, the range, range rate and acceleration observables will be corrupted by instrumental noise sources that cause offsets and drifts as a function of time, such as the noise due to the ultrastable oscillators (USOs). As is well known, unless such effects are properly modeled in the solution for gravity coefficients, these noise sources can map into large errors in estimated coefficients. The purpose of this appendix is not to promote a particular approach for modeling instrumental drifts but to illustrate concepts and potential complications, from a frequency-domain perspective, with regard to the potential mapping of low-frequency noise (i.e., drifts) into estimated parameters. Frequency-domain mapping is illustrated here with heuristic examples consisting of fits to estimate the amplitude of selected constant-frequency tones along with instrumental models comprised of polynomials. Even though actual GRACE partials are more complicated (e.g., with orbital variations, slowly changing tone frequencies and different tones with various relative tone amplitudes), the examples presented here can provide useful insights and crude guidelines. Because of the greater complexity of actual partials and fits, however, optimization of the instrumental model for actual fits requires a more complete approach. Even though the frequency-domain technique is applied here only to polynomial instrumental models, it can be modified to analyze other instrumental models with regard to their effectiveness in suppressing drifts.

Because of the wide variety of issues and properties implicitly present in a multiparameter solution for gravity coefficients, a relatively large number of heuristic examples are presented, illustrating various signal-tone combinations, update intervals, and instrumental polynomials. The examples suggest that, with appropriate polynomial modeling, instrumental drifts can be removed relatively effectively and instrumental parameters and gravity coefficients "separate" sufficiently to justify the "zero-correlation" approximation, at least for the hypothetical acceleration approach of this report. Further, the analysis supports the error-propagation approximation in Section 5 in which parameter noise is estimated by placing a bandpass of width $1/T_{\text{fit}}$ around partial tones. In Subappendix C.2.10, a simple example illustrates the fit-filter consequences of concurrently estimating a 1/rev orbit-parameter tone.

When a given model is used in estimation, one method for assessing its effectiveness is to compare resulting total noise (e.g., jitter determined from simulations) on each important estimated parameter with corresponding total noise produced by other models. Although such assessment can work in a relative sense for the given set of attempted cases, it provides little quantitative insight into internal properties of the fit. The frequency-domain approach provides a more detailed approach for assessing the possibility of further reducing the impact of random errors or error tones on a given estimated parameter, through better modeling. As explained below, this method is based on the presence of excessive mainlobes and/or sidelobes in the noise spectrum of the estimated parameter.

C.1 Removal a $f^{-\alpha}$ Singularity with a Polynomial Model

Random noise that has a power spectrum with a frequency dependence proportional to $f^{-\alpha}$, where α is real and positive, will cause drifts in time that can severely corrupt parameter estimation. One can see how such frequency dependence can cause damage by considering the frequency-domain error mapping predicted by the particular approximate fit filter in Eq. (5.11), which is generated by a fit model that only estimates the amplitude of the signal and therefore ignores the possibility of instrumental drifts. In this fit filter, the sinx/x functions have sidelobes that reach down to zero frequency. For a relatively flat noise spectrum, such sidelobes are not a problem since they are small and pick up relatively little noise power compared to the main lobe. However, if the noise spectrum varies as $f^{-\alpha}$ as f approaches zero, the noise components near zero frequency can be very large compared to the noise components within the main lobe and the noise power collected by the sinx/x sidelobes near $f = 0$ can be large compared to that collected by the main lobe. Thus, from the frequency-domain perspective, there is a need to modify the fit model so that the fit filter approaches zero faster than $G(f) \sim f^{\alpha}$ as $f \rightarrow 0$ in order to counteract the $f^{-\alpha}$ dependence of the noise.

A standard approach for coping with such drifts is to add a polynomial in time to the model, such as linear $(a + bt)$ or quadratic $(a + bt + ct^2)$ and estimate the polynomial coefficients along with the gravity coefficients. To gain insight with respect to the filtering effect of such terms, first consider the nature of the noise at low frequencies. Let the noise source under consideration have the following Fourier decomposition:

$$\eta(t) = \int_{-\infty}^{\infty} N(f)\, e^{2\pi i\, ft}\, df \tag{C.1}$$

for which the spectral noise density has the limit

$$S_{\eta}(f) \sim f^{-\alpha}, \quad \text{as } f \rightarrow 0 \tag{C.2}$$

as presumed above. For a given member from the ensemble of possible noise sequences, the time dependence associated with a given low frequency component, η_f, can be approximated over the fit interval as

$$\eta_f(t) \approx N(f)\left[1 + 2\pi i f t - (2\pi f t)^2\right], \quad \text{as } f \rightarrow 0 \tag{C.3}$$

Adding a constant offset term to the fit model will remove the constant term in this expansion, leaving the higher order terms. Consequently, at low frequencies, a fit that includes an estimated constant passes the second and higher terms in Eq. (C.3) and therefore has a frequency response (fit filter) that varies as

$$\sqrt{G(f)} \sim f, \quad \text{as } f \rightarrow 0 \tag{C.4}$$

Similarly, adding a linear function to the model will remove the first two terms in Eq. (C.3) causing the frequency response of the fit filter at low frequencies to vary as f^2. Thus, as the power of the polynomial increases, the low-frequency filtering becomes stronger.

When a polynomial with terms up to t^n is added to the estimated parameters, one can therefore show that the low-frequency dependence of the frequency response of the fit, in power units, decreases at least as fast as

$$G(f) \sim f^{2n+2}, \quad f \to 0 \tag{C.5}$$

Thus, when $f^{-\alpha}$ noise is present, the magnitude of α can be used to determine the order of the required instrumental polynomial. For example, the singularity at zero frequency is reduced to zero if $2n + 2 > \alpha$. Since the spectral noise density (in acceleration units) for the accelerometer is specified to vary approximately as f^{-1} (i.e., $\alpha = 1$) for low frequencies, a constant instrumental term ($n = 0$) or higher order is indicated when modeling this error source for the <u>acceleration</u> observable. In a second example, USO noise spectral density in the <u>range</u> observable varies as f^{-4} in power at low frequencies, in which case $n \geq 2$ is indicated.

The preceding analysis does not address the important issue of length of fit intervals. In a solution for gravity coefficients, as is well known, the total fit interval (e.g., 13 days) can be divided into a number of subintervals, each of length T_u, with a separate instrumental polynomial estimated for each subinterval but with gravity coefficients based on the whole interval. Subinterval length should be set to a value that is small in order to improve suppression of $f^{-\alpha}$ noise but should be sufficiently large to "separate" the partials for the instrumental parameters from the partials for orbit parameters and gravity coefficients. Since the frequencies of the important tones in the partials with respect to orbit and gravity coefficients are approximately once per revolution or higher, the minimum length for T_u is on the order of one revolution. That is, a minimum length of approximately 1 rev would cause a 1/rev tone in a partial to traverse at least one cycle over a subinterval, and should therefore start to "separate" that partial from an instrumental model that is constant, linear or quadratic over that same interval. A 2-rev update interval would more completely "separate" a 1/rev tone from such instrumental models. An example of an update interval that is too short is a piecewise quadratic instrumental model with a 0.5-rev update interval. In this example, a quadratic over a half revolution can look too much like a half cycle of a 1/rev sinusoid in a partial. This and other heuristic examples presented in the following subappendices illustrate the importance of properly selecting the update interval for instrumental parameters.

C.2 Fit Filters in the Frequency Domain for Some Simple Examples

C.2.1 Frequency-Response Equations

Even though the "$2n + 2 > \alpha$" rule for polynomial selection suppresses the "singularity" at zero frequency associated with an $f^{-\alpha}$ noise power spectrum, fit-filter sidelobes can still reach down to very low frequencies if the instrumental-parameter update interval is not properly selected. Such low-frequency sidelobes can pick up noise components generated by a $f^{-\alpha}$ spectrum near zero frequency, components that can be substantial when α is large (e.g., $\alpha = 4$). Under certain conditions, this sidelobe noise power can increase total integrated noise on the estimated parameters to levels much greater than is necessary.

To investigate fit-filter sidelobes, it is useful to derive the frequency response ("fit filter") for estimated amplitude (gravity coefficient) based on an evenly weighted ("white-noise") least-squares fit. Estimated parameters can be obtained through use of the well-known expression

131

$$\hat{\mathbf{p}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y} \qquad (C.6)$$

where $\hat{\mathbf{p}}$ is the estimated parameter vector for N_p parameters given by

$$\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \; .....\hat{p}_{N_p})^T , \qquad (C.7)$$

$\mathbf{y}$ is the observable vector for N_o observables over the fit interval, T_{fit}, given by

$$\mathbf{y} \equiv (y_1, y_2, \; ....y_{N_o})^T , \qquad (C.8)$$

and $\mathbf{A}$ is the associated partials matrix given by

$$A_{ij} \equiv \frac{\partial y_i}{\partial p_j} , \; i = 1, N_o; \; j = 1, N_p \qquad (C.9)$$

To some approximation, $\mathbf{y}$ in this analysis can be thought of as the range, range rate or acceleration.

The frequency response for random noise at a given frequency, f, can be obtained by substituting, in turn, the following two quadrature sinusoidal "observables" in Eq. (C.6):

$$\mathbf{y} \equiv (\cos(2\pi f t_1), \cos(2\pi f t_2), \; ....\cos(2\pi f t_{N_o}))^T \qquad (C.10)$$

and

$$\mathbf{y} \equiv (\sin(2\pi f t_1), \sin(2\pi f t_2), \; ....\sin(2\pi f t_{N_o}))^T \qquad (C.11)$$

where the t_i are the timetags of the observations (e.g., at 10-s intervals). For a given estimated parameter (i.e., a given element in the $\hat{\mathbf{p}}$ vector), response at each given frequency is computed as the RSS of the outputs for these two quadrature inputs. (A quadrature input is appropriate for random noise since, in effect, both quadrature components are present in that case. For a pure tone error, however, a sinusoidal time sequence with phase appropriate for the actual tone would be substituted in Eq. (C.6).) Note that both input amplitudes are 1.0 so that output at a given frequency can be regarded as "gain" at that frequency.

In the case of white input noise on $\mathbf{y}$, with variance σ_y^2, the error in an estimated parameter (often referred to as the "formal error") can be computed as the square root of the appropriate diagonal element of $(\mathbf{A}^T\mathbf{A})^{-1}\sigma_y^2$. When properly normalized by the formal errors, the off-diagonal elements of $(\mathbf{A}^T\mathbf{A})^{-1}$ predict the correlations between estimated parameters, again in the case of white input noise. Formal errors and correlations computed in this fashion can be unreliable when the actual observable noise is colored. For substantially colored input noise, as seen below for specific examples, actual parameter errors can be quite different from formal errors.

As a crude measure of the importance of correlations between estimated parameters, "formal-error amplification" for a given estimated parameter is defined here as the ratio of two errors: the total error in the parameter when the parameter is estimated alone and the total error when the parameter is estimated along with other parameters, <u>given a flat input noise spectrum</u>. This ratio can be computed as the square root of the ratio of appropriate diagonal elements in $(\mathbf{A}^T\mathbf{A})^{-1}$ for said two estimations. Alternately, the ratio can be computed as the ratio of total errors for the given parameter, where each total error is computed as the square root of the integral of the power frequency response for a flat noise spectrum, over all frequencies. Even though formal-error amplification can be used in some sense as an indicator of the importance of correlations between parameters, it should not be assigned too much quantitative significance in cases for which actual input noise is colored.

To illustrate the use of polynomial models in handling a $f^{-\alpha}$ spectral noise density, results are now presented for a number of heuristic examples

C.2.2 Single Tone and No Instrumental Model

As a first, simple illustration, Fig. C-1 presents the frequency-domain fit filter for a fit that estimates the amplitude, A, of a sine tone with a constant frequency of 2/rev. In this first example, there is no instrumental model so that only one parameter, A, is estimated. The purpose of this example is to introduce some basic features of fit filters and to illustrate the aforementioned singularity at zero frequency, in the case of $f^{-\alpha}$ noise, when no instrumental parameters are estimated. The lowest curve in the figure is the fit filter for A (frequency response in "voltage") computed on the basis of the unit-amplitude quadrature inputs in Eqs. (C.10) and (C.11), as explained above. Note that the frequency axis is in "cycles-per-rev" units where one revolution is approximately 5500 s.

Before discussing the singularity issue, some basic features of fit filters for amplitude should be pointed out. In Fig. C-1, the fit filter for A consists of a mainlobe centered at the "signal" tone frequency of 2/rev, with first nulls at $1/T_{fit}$ and surrounded by sidelobes that monotonically decrease in magnitude as separation from the mainlobe increases. (Asymmetry in sidelobe magnitude about the mainlobe can be viewed as a consequence of the influence of a "negative-frequency" component of the fit filter centered at $f = -2$/rev.) Note that filter magnitude at a frequency of 2/rev is 1.0, which means a sine tone at 2/rev will be passed with unit gain, as required. (That is, if the input sine tone has amplitude A, the output of the fit will be A.) As suggested by this example, the fit filter for a given amplitude parameter generally exhibits mainlobes that are centered at the frequencies of the tones found in the partial for that amplitude. As illustrated below, however, relative magnitudes of the mainlobes can be substantially different than relative amplitudes of the tones in the partial and some (potential) mainlobes can even be suppressed.

Except for excessive sidelobes, as discussed below, sidelobe magnitude at a given frequency in an amplitude fit filter generally decreases in proportion to $1/T_{fit}$ as fit interval, T_{fit}, increases. Thus, in this example, sidelobe magnitude at a given frequency would decrease by approximately an additional factor of 36 relative to the mainlobe if the fit interval were increased from 5.75 revs to a more realistic gravity-solution interval of 13 days (200 revs). Sidelobe width, which is approximately equal to $1/T_{fit}$, also decreases as fit interval increases. Thus, for a more realistic fit interval of 13 days, the fit filter of Fig. C-1 would consist of a very narrow ($\approx 1\ \mu$Hz ≈ 0.005/rev) mainlobe spike at 2/rev, with a peak value of 1.0 that towers above small, very narrow sidelobes (e.g., the mainlobe would be on the order of 300x higher than the sidelobes at 1/rev). This example illustrates a point to remember in considering the following plots: The important features in the plots are generally the mainlobes and excessive sidelobes. Even though the "normal" sidelobes surrounding the mainlobes are relatively large in the plots, they would decrease dramatically in magnitude relative to the mainlobes if integration time were increased to a more realistic value of, say, 13 days.

133

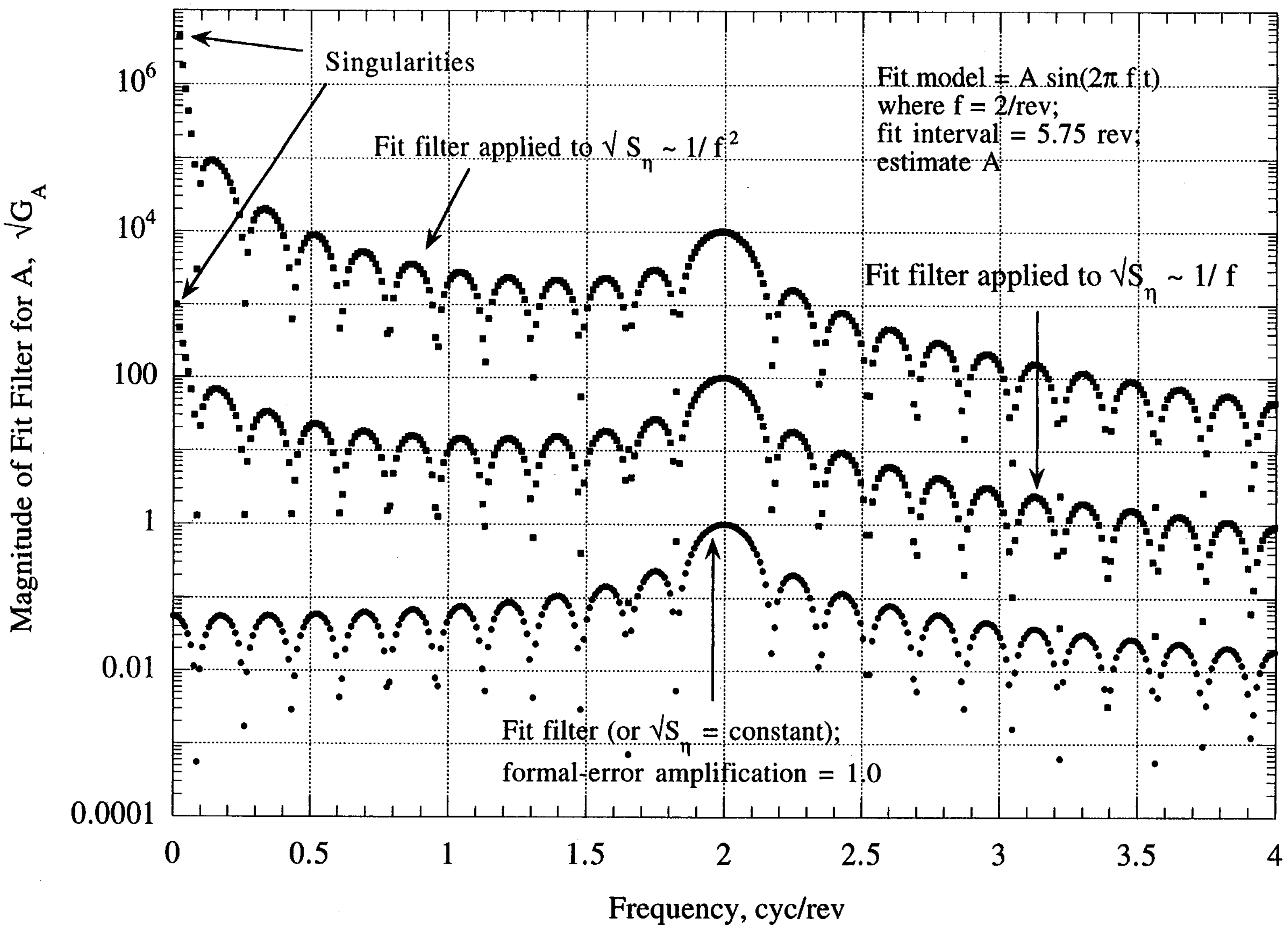

Figure C-1 Example of a Fit Filter for the Amplitude of a 2/rev Sine Function

Except for a multiplicative scale factor, the bottom curve in Fig. C-1 can be regarded as the output for a flat input noise spectrum ($\sqrt{S_\eta}$ = constant). By definition, formal-error amplification for A is 1.0. To better demonstrate singularity, fit interval and tone frequency have been selected in this plot to place a sidelobe maximum at zero frequency. For illustration, Fig. C-1 also shows the noise spectra on the fit output for A for two input noise spectra: $\sqrt{S_\eta} \sim 1/f$ (i.e., $\alpha = 2$) and $\sqrt{S_\eta} \sim 1/f^2$ (i.e., $\alpha = 4$) in "voltage" units. These curves are computed by multiplying $\sqrt{S_\eta}$ by the fit filter, with a somewhat arbitrary magnitude scaling applied to separate the curves for plotting purposes. Since no instrumental parameters are estimated, the aforementioned singularity at zero frequency is present in both of these curves. (Divergence to infinity as frequency approaches zero would be more evident if smaller step size in frequency had been used.) These divergences graphically illustrate the need for better instrumental modeling.

C.2.3 Single Tone Plus Quadratic Instrumental Model

Fig. C-2 illustrates the suppression of the singularities in Fig. C-1 that can be obtained by estimating a quadratic instrumental function in addition to the amplitude of a 2/rev sinusoid. In this example, only one quadratic instrumental function is estimated, leading to a total of four estimated parameters, for a fit interval of 6 revs. The lowest curve in the figure is the fit filter for the parameter A (again frequency response is in "voltage"). As one would expect, the width of the mainlobe for this example and the following examples is still $\pm 1/T_{fit}$ to first nulls. Fit interval and tone frequency have been selected in this example to provide a worst-case example for the low-frequency sidelobes of the fit filter. (That is, "positive-frequency" and "negative-frequency" components of the fit filter combine constructively to form the sidelobes.) For a flat input noise spectrum, the total noise on A (e.g., obtained from integration of power over frequency in the bottom curve) turns out to be essentially the same as that for a fit to A alone in Fig. C-1. That is, formal-error amplification is approximately 1.0, indicating that concurrent estimation of the three additional instrumental parameters over the whole fit interval does not degrade the error on A, relative to estimating A alone, for a <u>flat</u> spectrum.

Near zero frequency, fit filter magnitude in Fig. (C-2) (lowest curve) varies as f^3 as a result of the quadratic instrumental model as explained above. As shown in the two upper curves in Fig. (C-2) and as anticipated above, this f^3 dependence suppresses the singularities at zero frequency found in Fig C-1 for 1/f and $1/f^2$ input noise spectra. For the 1/f spectrum, the sidelobes are relatively small: the largest sidelobe is approximately 60% of the mainlobe, indicating that the increase in total integrated noise caused by the lower sidelobes is relatively small (< 20%) but could be improved.

Filtering for the $1/f^2$ case (i.e., $\alpha = 4$), the top curve in Fig. C-2, is unacceptable. Even though the fit filter suppresses the $f^{-\alpha}$ singularity at $f = 0$ as expected, the first sidelobe above zero frequency is much larger (by approximately a factor of 7.8) than the mainlobe at 2/rev ($\approx$ 0.36 mHz). Consequently, total (integrated) noise on the estimated value for A is approximately eightfold larger (64x in power) than the noise under the mainlobe. If this excessive sidelobe noise could be suppressed, the error in A could be greatly reduced.

This heuristic example illustrates how a whole-interval quadratic instrumental model can suppress the singularity at zero frequency caused by $f^{-\alpha}$ noise, as anticipated in Subappendix C.1. However, excessive sidelobes at low frequencies can still substantially degrade estimated amplitude. This type of excessive sidelobe, the type caused by fit-filter sidelobes being too large at low (<< 1/rev) frequencies (to be called Type-I excessive sidelobes), can be further suppressed by using a piecewise instrumental model, as shown in the following subappendices.

139

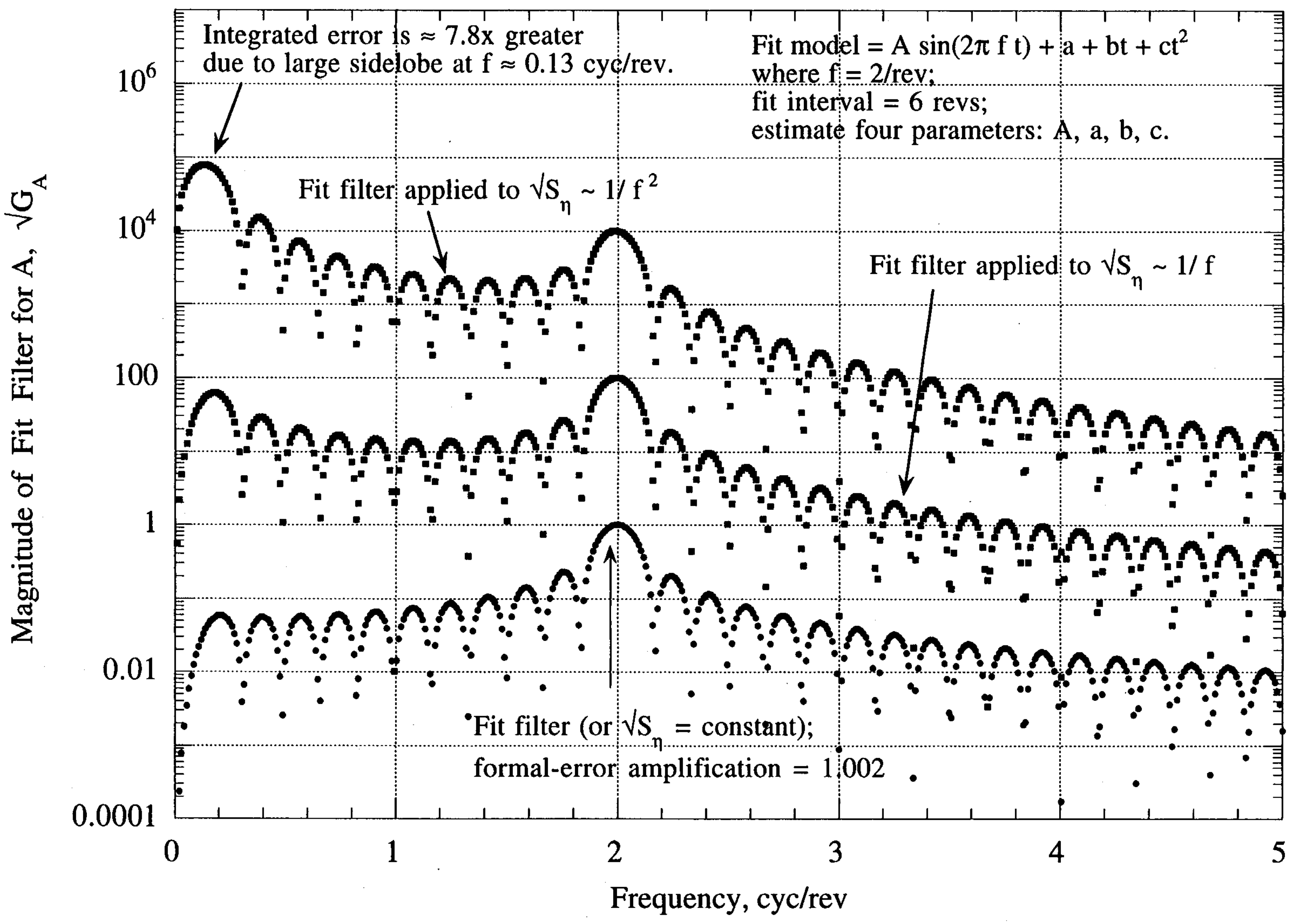

Figure C-2 Fit Filter for the Amplitude of a 2/rev Sine Function when Estimated along with a Quadratic Instrumental Function

C.2.4 Single Tone Plus Piecewise Quadratic Instrumental Model

Fig. C-3, in which a short update interval is applied to the case in Fig. C-2, illustrates the impressive reduction in excessive sidelobe noise (Type I) that can be obtained by reducing the quadratic-fit interval from the whole interval (6 revs) to 1 rev. As seen in the fit filter represented by the lowest curve in Fig. C-3, the shorter update interval greatly reduces the sidelobe magnitude (Type I) relative to Fig. C-2 at frequencies below f = 1/rev, as one would expect (see next paragraph). Note, however, that including a piecewise-quadratic model introduces a new "resonance sidelobe" at 1/rev (to be called a Type II excessive sidelobe) caused by a strong interplay between the 2/rev "signal" and the piecewise quadratic as the fit combines them in an attempt to mimic a 1/rev signature. Other, smaller resonance sidelobes are introduced at 3, 4, 5/rev, etc. In spite of the increase caused by these resonance sidelobes, total noise under all lobes for the $1/f^2$ case is smaller than in Fig. C-2 by approximately a factor of 4.5.

Suppression of Type-I sidelobes provided by the preceding shorter update interval can be semiquantitatively explained as follows. Consider fitting a quadratic over an update interval, T_u, to a noise sinusoid of frequency, f, that traverses one half cycle over the interval (i.e., $f = 0.5/T_u$). For such a frequency, the shape of the noise sinusoid over the interval has, to first approximation, a linear or quadratic appearance and the quadratic can do a fairly good job of removing it. The quadratic can do an even better job of removing noise sinusoids at lower frequencies, improving as frequency decreases. At higher noise frequencies on the other hand, the quadratic becomes less effective as frequency increases. For example, a quadratic is less effective at removing noise with a frequency that traverses one cycle over an update interval ($f = 1/T_u$). These examples indicate that a quadratic with an update interval, T_u, can very effectively suppress noise components with frequencies on the order of $0.5/T_u$ or less, with improving suppression as frequency decreases.

As indicated by the top curve in Fig. C-3, the total error in A could be further reduced in the $1/f^2$ case (by approximately a factor of 1.6) if the resonance sidelobe (Type II) at 1/rev could also be suppressed. Increasing fit-interval length does not reduce this type of excessive sidelobe. Another possibility is to make the update-interval length less commensurate with 2/rev. However, when other update-interval lengths (e.g., 17/16, 9/8, and 5/4 rev) are used for the 2/rev signal over 5 update intervals, a similar, though slightly smaller, resonance sidelobe appears in each case at approximately 1/rev. (See Subappendix C.2.9 for improvement gained by varying update-interval length over a 13-day fit interval.) Formal-error amplification for this example is approximately 1.09, indicating that a minor loss in accuracy is suffered in the case of a flat input noise spectrum. Overall, a piecewise quadratic with 1/rev updates does a fair job of suppressing $1/f^2$ noise, given a 2/rev signal.

Figs. C-4 and C-5 change the update interval for the preceding example of a 2/rev signal to 2 revs and 3 revs, respectively. As one would expect, the $1/f^2$ random noise mapped to estimated amplitude increases as update interval increases in length. Figs. C-4 and C-5 show what happens in the frequency domain to cause this increase in error. As the length of the update interval increases, the location of the largest excessive resonance sidelobe shifts to lower frequency (from 1/rev for 1 rev updates to 0.33/rev for 3 rev updates) and its fit-filter magnitude decreases (from 0.36 for 1 rev updates to 0.106 for 3 rev updates). As indicated by the uppermost curve, this decrease in fit-filter magnitude is not sufficient to compensate for the larger noise at lower frequencies in the case of $1/f^2$ noise. For 2 rev updates, the excessive-sidelobe noise magnitude is 2.5 times the mainlobe magnitude for $1/f^2$ noise compared to 1.4 times for 1 rev updates. Thus, for this special example of a 2/rev signal, 2 rev updates are somewhat worse than 1 rev updates in the case of $1/f^2$ random noise (approximately 1.6 times worse in total noise).

131

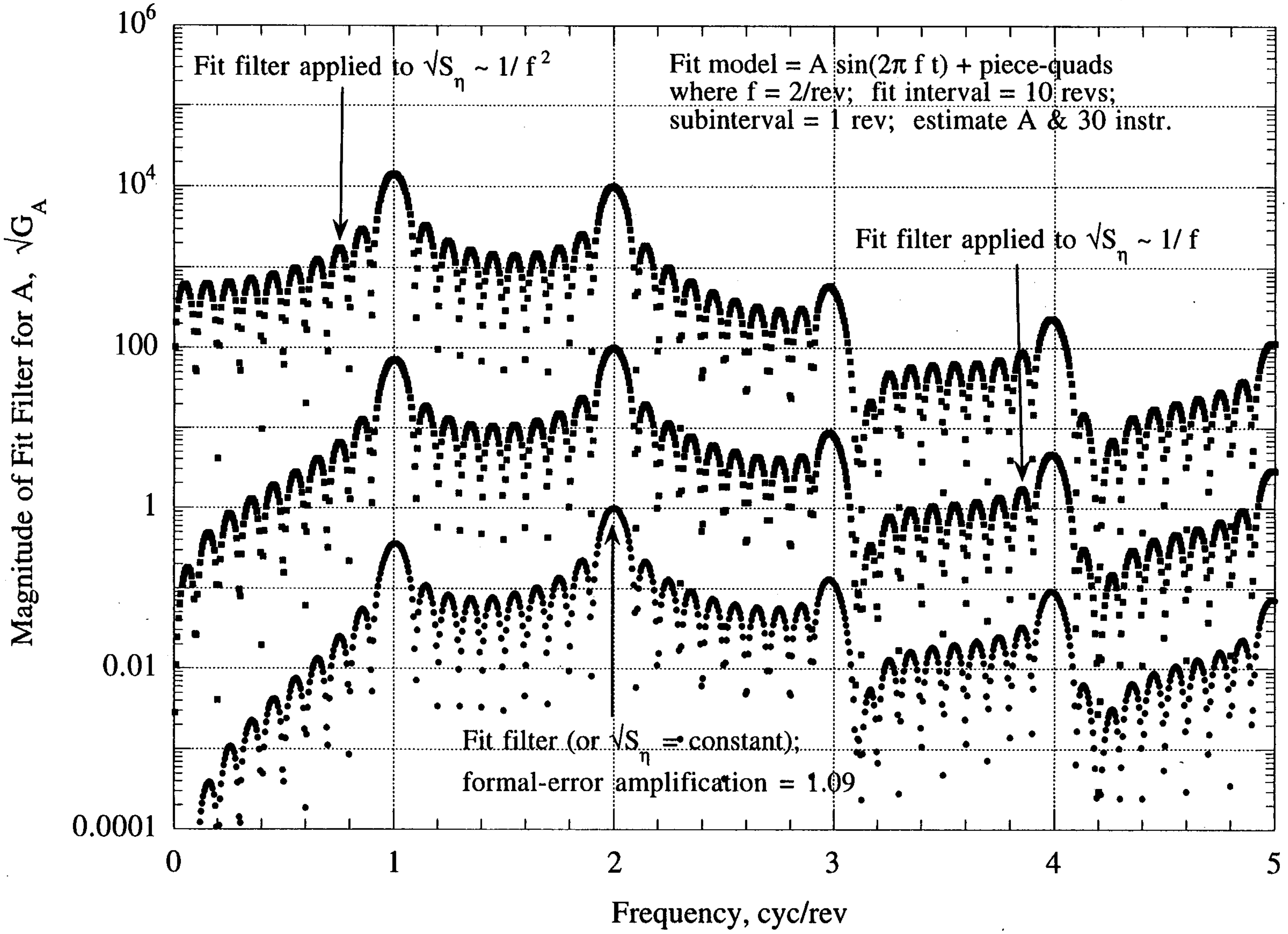

Figure C-3 Fit Filter for the Amplitude of a 2/rev Sine Function when Estimated along with a 1-rev-piecewise Quadratic Function

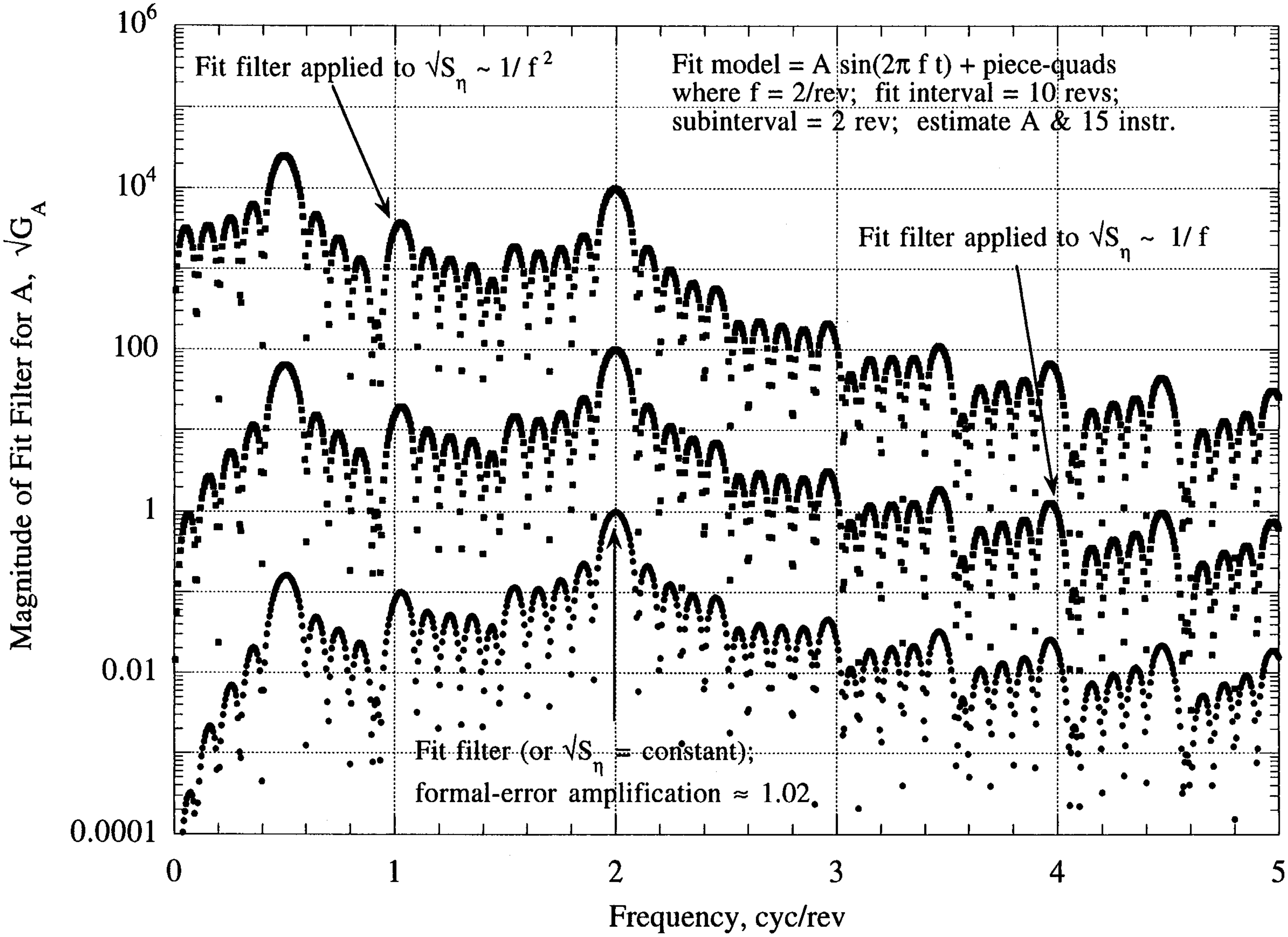

Figure C-4 Fit Filter for the Amplitude of a 2/rev Sine Function when Estimated along with a 2-rev-piecewise Quadratic Function

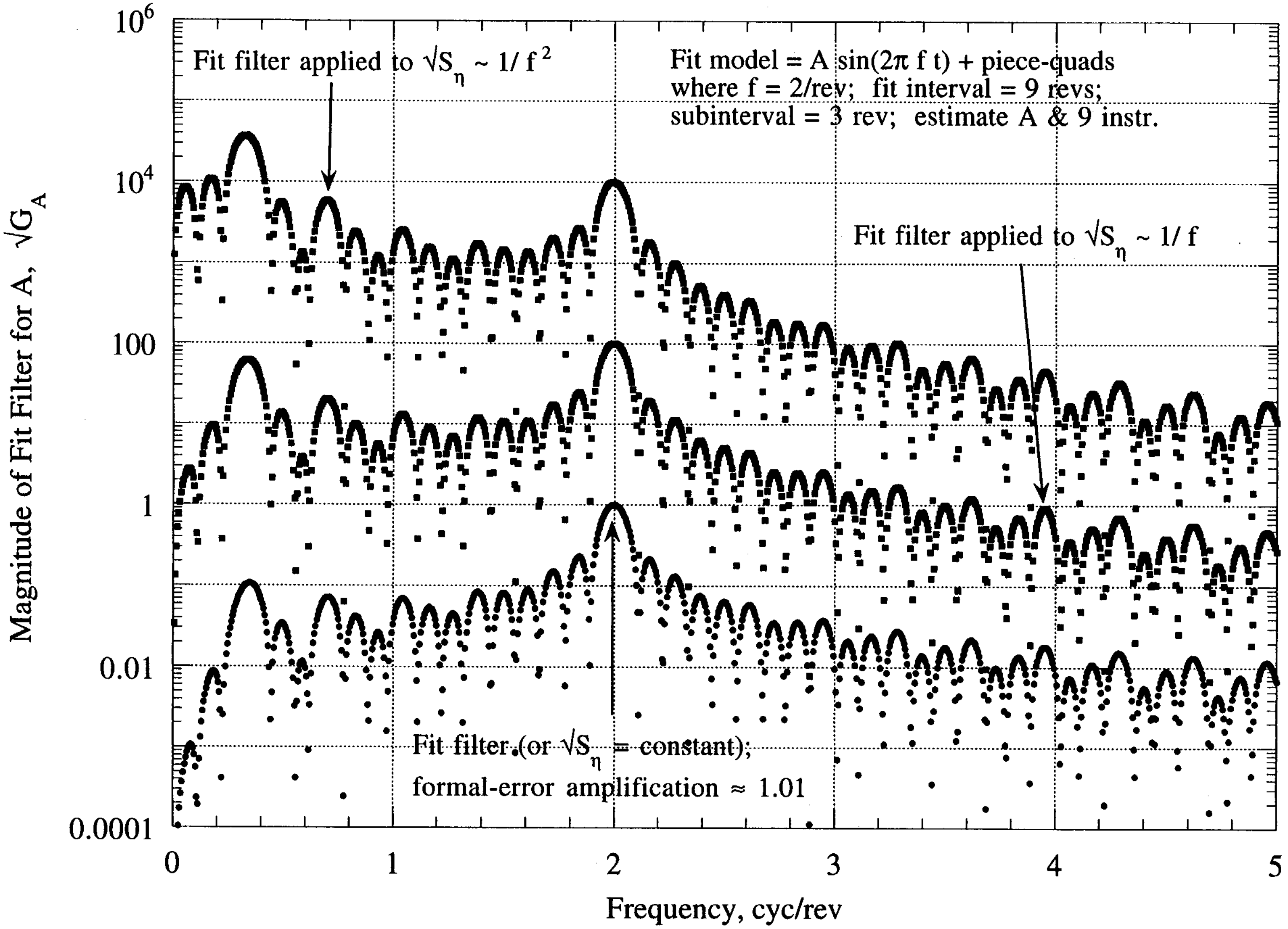

Figure C-5 Fit Filter for the Amplitude of a 2/rev Sine Function when Estimated along with a 3-rev-piecewise Quadratic Function

For 1/f random noise (middle curve), the largest resonance sidelobe is approximately the same magnitude in Figs. C-3, C-4, and C-5 so that total noise would be about the same for the three update intervals (i.e., approximately a factor of 1.2 increase in total noise due to the excessive sidelobe for 1, 2, or 3 rev updates). However, since there are more excessive sidelobes for the longer update intervals, 1 rev is a better choice in the case of a 2/rev signal corrupted by 1/f random noise. For the same reason, 1 rev update interval is also a better choice for white noise. Thus, for white, 1/f and $1/f^2$ random noise, a 1 rev update interval is preferable to a 2 rev interval for a 2/rev signal, particularly for $1/f^2$ noise.

An excessively short update interval can cause an unnecessary increase in fit noise. Fig. C-6 illustrates the fit filter for a model consisting of signal at 1/rev plus a piecewise quadratic with 0.5-rev update interval. In this case, formal-error amplification is approximately 41, suggesting that, in the case of a flat input noise spectrum, the total noise mapped in the case of a 0.5-rev update interval is far greater than the noise that would be mapped with a better update interval. In this example, the excess noise is caused by resonance sidelobes (mainly at 3/rev) above the signal frequency of 1/rev. Figs. C-7 and C-8 present the fit filters for the same 1/rev signal but with the update interval increased from 0.5 rev to 1 rev and 2 revs, respectively. In these examples, formal-error amplifications are much smaller at 1.6 and 1.09, respectively. Thus, a 2 rev update interval would be somewhat better for a 1/rev signal with white noise. For 1/f random noise, a 2 rev update interval increases the error by approximately 25% while a 1-rev update interval increases the error by approximately 8%. Thus, for 1/f random noise, a 1-rev update interval would be marginally superior. As noted in the figures, the 1-rev-update example increases $1/f^2$ noise by less than 3%, while 2-rev example increases $1/f^2$ noise by approximately a factor of 1.7. Thus, a 1 rev update interval would be somewhat better for $1/f^2$ noise. In these examples, the substantial sidelobes above 1/rev for the 1-rev updates in Fig. C-8 could be a problem if sufficiently large error tones were located at those frequencies (e.g., at 2/rev, 3/rev, etc.). In this case, a longer update interval is indicated.

These results suggest a fixed update interval in the range of 1 rev to 2 revs, depending on error characteristics, can provide reasonably good results, but also indicate a better instrumental model should be possible, as discussed in Subappendix C.2.9.

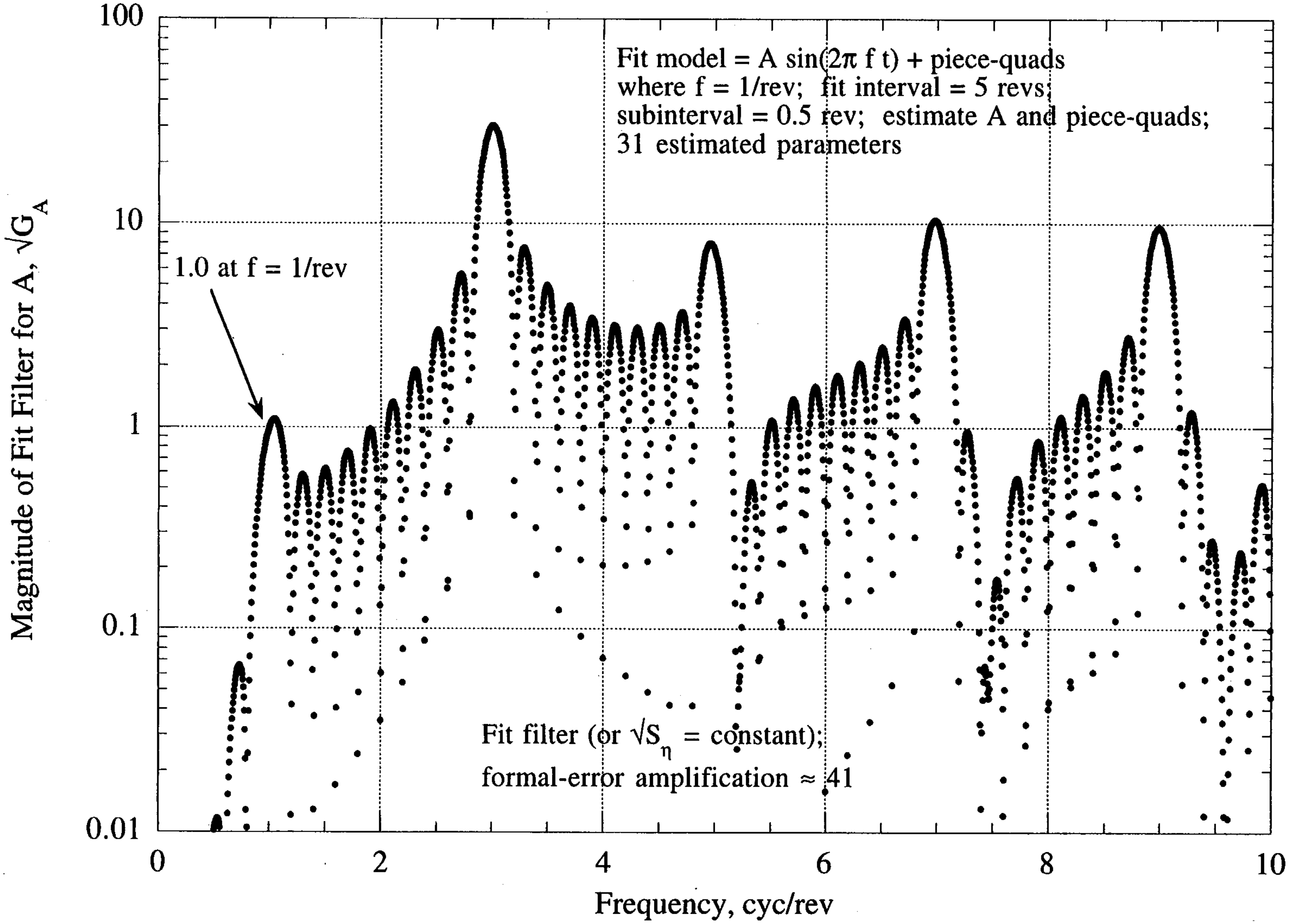

Figure C-6 Fit Filter for the Amplitude of a 1/rev Sine Function when Estimated along with a 0.5-rev-piecewise Quadratic Function

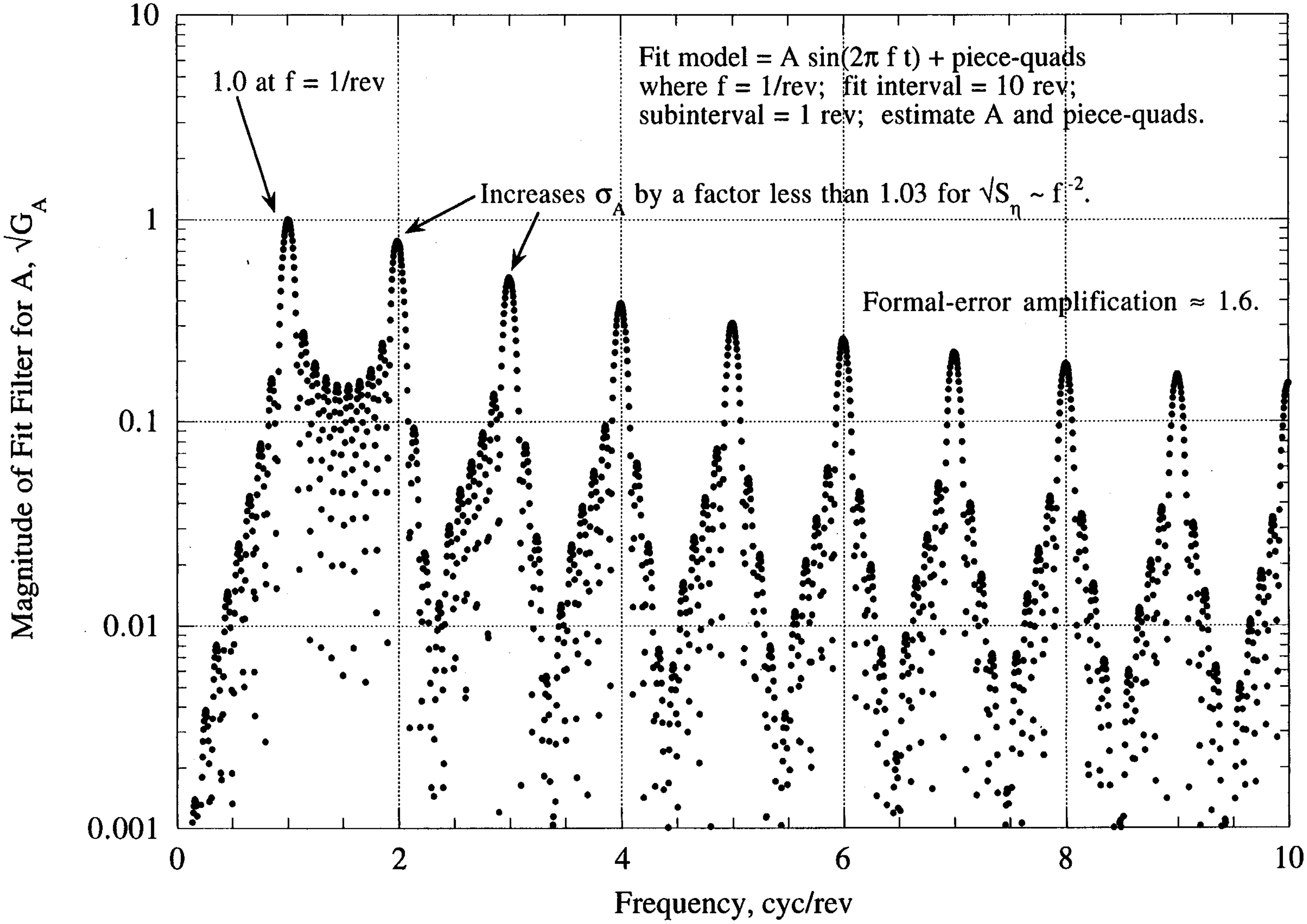

Figure C-7 Fit Filter for the Amplitude of a 1/rev Sine Function when Estimated along with a 1-rev-piecewise Quadratic Function

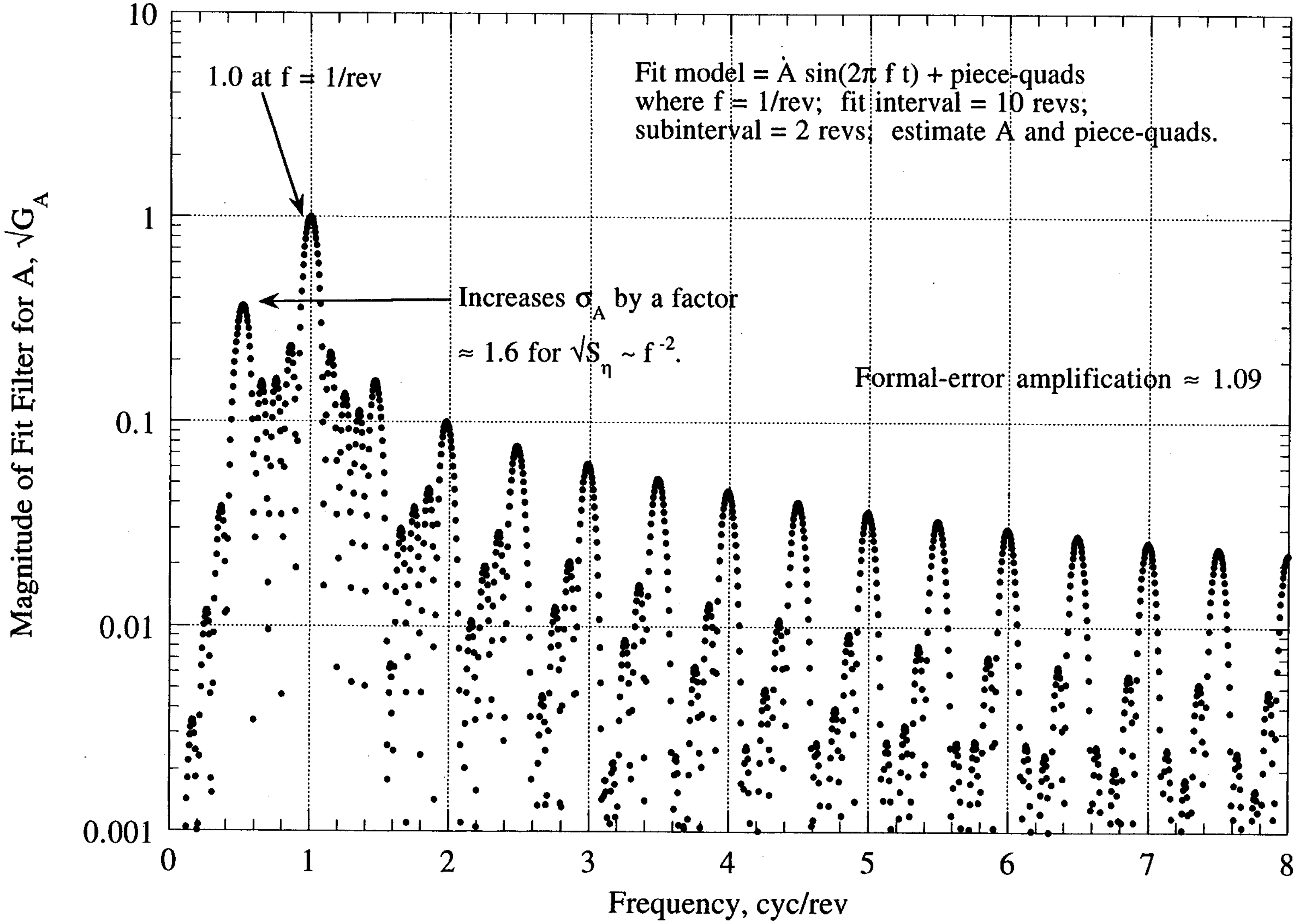

Figure C-8 Fit Filter for the Amplitude of a 1/rev Sine Function when Estimated along with a 2-rev-piecewise Quadratic Function

C.2.5 Two-Tone Signal

To illustrate fits involving a signal comprised of two tones, Figs. C-9 and C-10 present fit filters for the overall amplitude of a signal comprised of equal-amplitude cosine tones at 1/rev and 3/rev, where the fit model for Fig. C-9 includes no instrumental terms and the fit for Fig. C-10 includes a piecewise-quadratic instrumental model with 1-rev updates. As one might anticipate, the fit filter in Fig. C-9 has two mainlobes of equal magnitude (0.5) centered at the two signal tones. By definition, the formal-error amplification for this example is 1.0. In Fig. C-10, inclusion of the piecewise-quadratic instrumental model with 1-rev updates reduces the magnitude of the 1/rev mainlobe from 0.5 to approximately 0.03 and increases the magnitude of the 3/rev mainlobe from 0.5 to approximately 1.03. (The 1/rev mainlobe is reduced in magnitude because a cosine at 1/rev looks somewhat like a quadratic over 1 rev. The analogous case with a sum of two sine functions has a less drastic reduction of the 1/rev mainlobe magnitude, i.e., to 0.2 rather than 0.03 because a sine at 1/rev looks less like a quadratic.) This reduction in the 1/rev mainlobe helps substantially in decreasing total noise in A in the case of $1/f^2$ noise. Note that a "resonance" sidelobe appears in Fig. C-10 at a frequency of approximately 2/rev ($\approx$ 0.37 mHz) with a magnitude of approximately 0.31. This 2/rev resonance sidelobe increases total noise by approximately a factor of 1.2 in the case of $1/f^2$ noise. Otherwise, noise is contributed primarily by the mainlobe at 3/rev. In contrast, as indicated above, the error in A for the filter in Fig. C-9 is formally infinity for $1/f^2$ noise since the singularity at zero frequency is not suppressed. Formal-error amplification in Fig. C-10 increases to approximately 1.5, mainly as a result of the increased magnitude of the mainlobe at 3/rev relative to Fig. C-9. Thus, the improvement provided by a 1 rev piecewise-quadratic model for $1/f^2$ noise is dramatic, restricting noise primarily to a $1/T_{fit}$ bandpass around 3/rev, while the degradation for white noise is only a factor of 1.5 relative to no instrumental model.

These results suggest that a piecewise quadratic model with 1-rev updates would handle both white noise and $1/f^2$ noise reasonably well for the indicated two-tone signal, but that some room for improvement remains. Further, Fig. C-10 illustrates a significant effect: the relative amplitude of tones found in the partial for a given parameter are not necessarily preserved in the fit filter for that parameter; in this example, the 1/rev tone is highly suppressed in the fit filter for amplitude by the piecewise quadratic instrumental model.

146

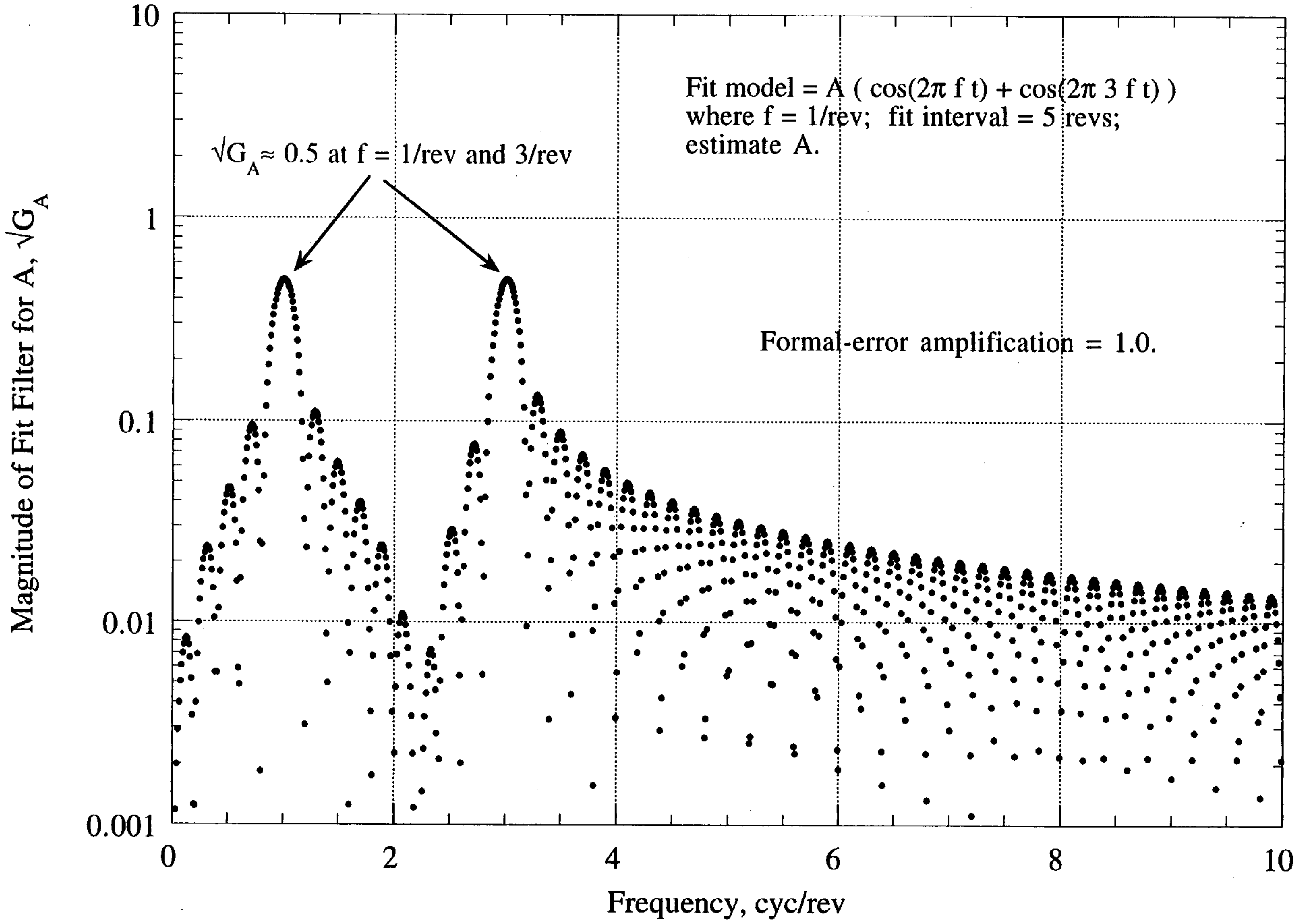

Figure C-9 Fit Filter for the Amplitude of the Sum of Two Cosine Functions

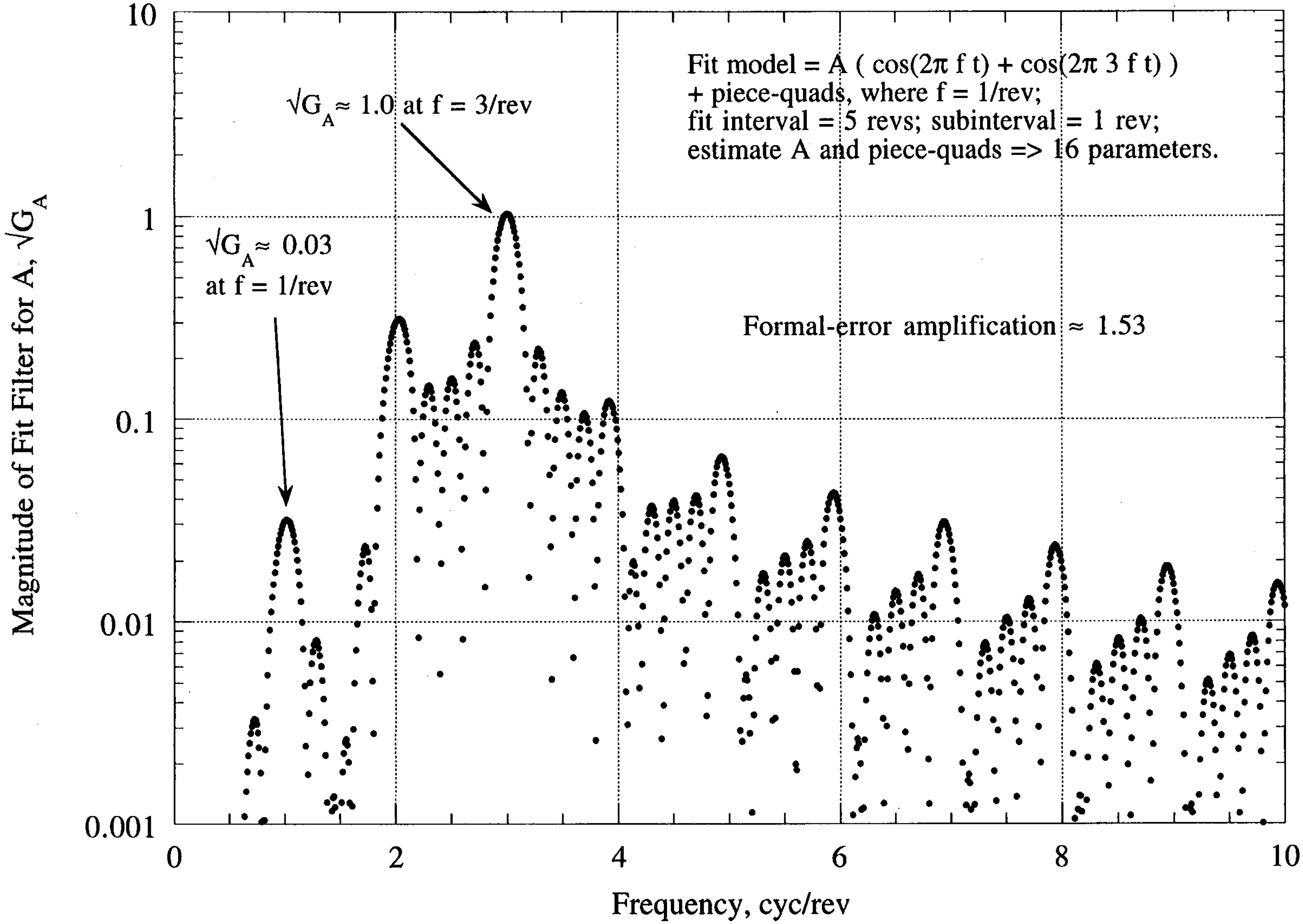

Figure C-10 Fit Filter for the Amplitude of a Sum of Two Cosines when Estimated along with a 1-rev-piecewise Quadratic Function

C.2.6 Multitone Signal

As examples of a multitone signal, Figs. C-11 and C-12 present fit filters for a signal consisting of a sum of cosines with frequencies equal to the first five odd multiples of 1/rev, where Fig. C-11 includes no instrumental terms and Fig. C-12 includes a piecewise-quadratic instrumental model with 1-rev updates. As one would expect, the fit filter in Fig. C-11 has five mainlobes of equal magnitude (≈ 0.2) centered at the five input tone frequencies. By definition, the formal-error amplification for this example is 1.0. In Fig. C-12, inclusion of the piecewise-quadratic instrumental model reduces the magnitude of the 1/rev mainlobe approximately from 0.2 to 0.03 and slightly increases the magnitude of other mainlobes approximately from 0.2 up to 0.24 to 0.27. Formal-error amplification is only 1.16. Comments similar to those for Figs. C-9 and C-10 can be made for Figs. C-11 and C-12. In Fig. C-12, the resonance sidelobe at 2/rev increases the total noise by a factor of approximately 1.2 in the case of $1/f^2$ noise. These results suggest that a piecewise quadratic model with 1-rev updates would handle white input noise and $1/f^2$ noise reasonably well for multitone partials.

C.2.7 A Sidelobe Method for Assessing Excess Noise

The preceding examples suggest a method for assessing whether a given instrumental model might be improved when solving for gravity coefficients: if possible, the sidelobes (Type I or Type II) in the error spectrum for a given coefficient (in response to the expected noise spectrum) should make a minor contribution to total integrated noise. Total noise in the estimated parameter should be contributed primarily by narrow mainlobes centered at tones in the partial for that parameter and not by the sidelobes. With regard to mainlobe magnitudes, however, the preceding examples indicate that inclusion of the instrumental model can suppress mainlobes in the fit filter that correspond to low-frequency tones (e.g., < 1/rev for 1 rev updates) in the partial. In addition, as indicated below in Subappendix C.2.11, magnitudes of mainlobes can be amplified as a result of correlations between estimated parameters.

In the final analysis, of course, the best model is the one that produces the least integrated noise for the expected input noise spectrum. Noise on an estimated parameter can be estimated either analytically by using methods such as those outlined above, or by numerical simulation based on random numbers generated with the expected spectral characteristics. Since GRACE involves simultaneous estimation of many orbital, gravitational and instrumental parameters, some with multiple tones in associated partials, it may not be possible to simultaneously "optimize" the instrumental model for all parameters. However, the preceding results suggest it is likely that all gravity coefficients can be estimated without excessive increase in error due to excessive sidelobe noise or due to correlations induced by the instrumental model.

C.2.8 Modeling $1/\sqrt{f}$ Noise

The examples in this subappendix pertain to the hypothetical acceleration simulations in Section 6 where fits are applied to simulated noise to obtain gravity coefficient errors, based on an instrumental model consisting of one constant for the whole 13-day fit interval. As a simple related example, Fig. C-13 presents results for a fit model consisting of a sine function at 1/rev plus one constant instrumental parameter, for a fit interval of 5 revs, where the two estimated parameters are the amplitude, A, of the sine function and the instrumental constant. The lower curve plots the fit filter for A while the upper curve plots the corresponding error spectrum for A given $1/\sqrt{f}$ input noise. For the $1/\sqrt{f}$ curve, the largest sidelobe is the lowest-frequency sidelobe (Type I) at approximately 0.1/rev, with a magnitude that is approximately 43% of the mainlobe. This sidelobe increases total noise by approximately 9% in this example.

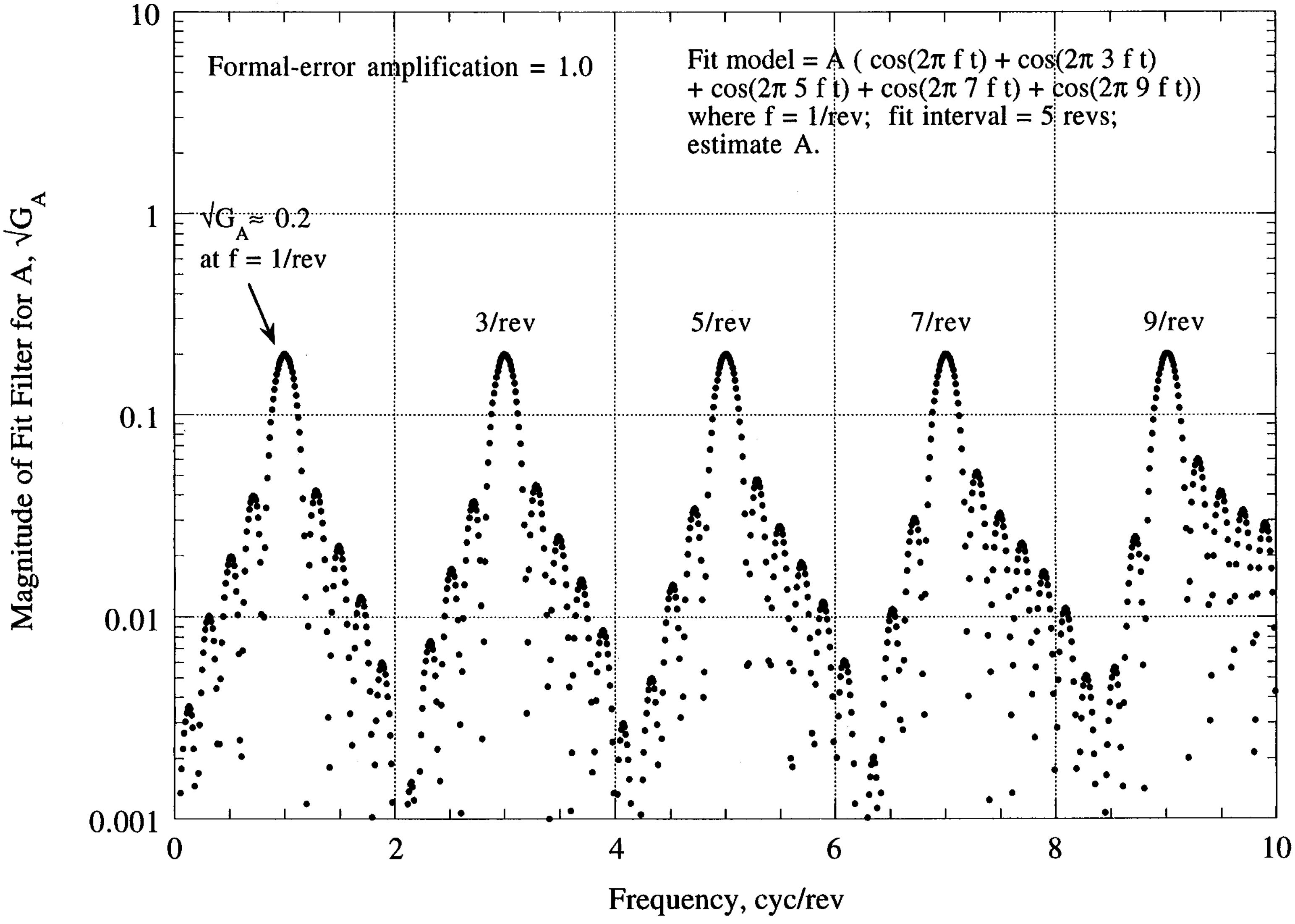

Figure C-11 Fit Filter for the Amplitude of a Sum of Five Cosines

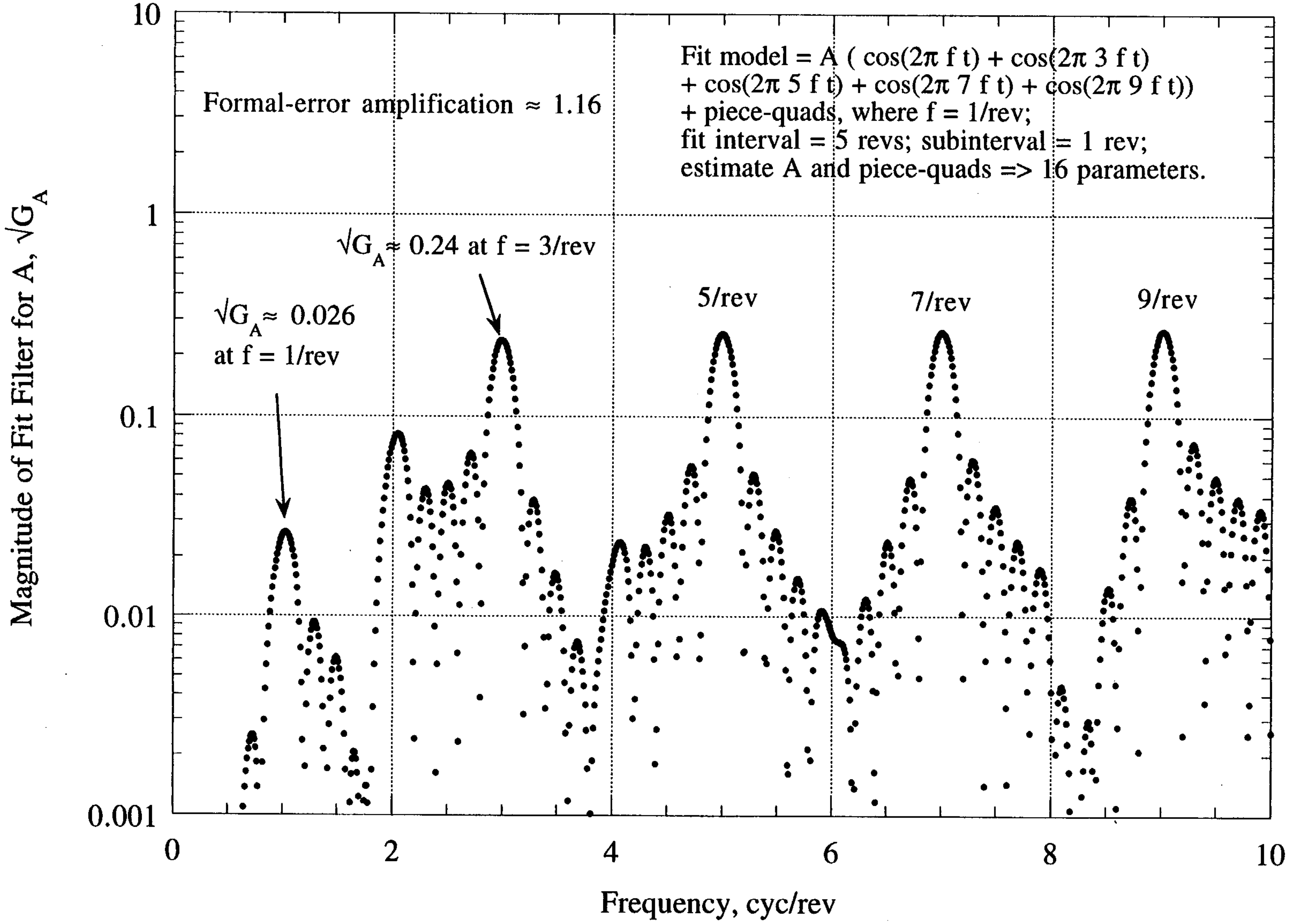

Figure C-12 Fit Filter for the Amplitude of a Sum of Five Cosines when Estimated along with a 1-rev-piecewise Quadratic Function

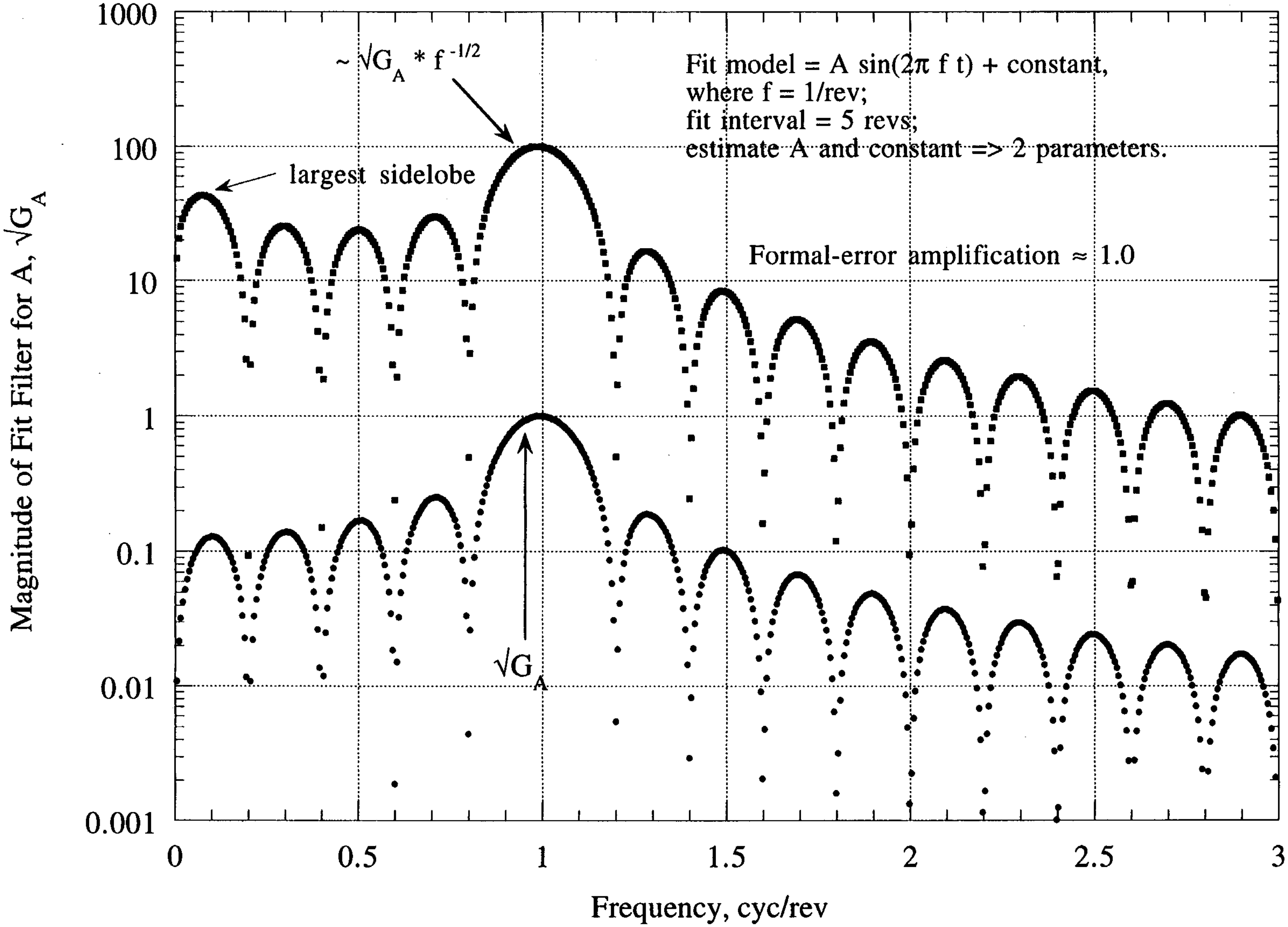

Figure C-13 Fit Filter for the Amplitude of a 1/rev Sine Function when Estimated along with a Constant Over 5 Revolutions

For longer fit intervals, frequency placement and the magnitude of this lowest-frequency sidelobe in the fit filter (lower curve) decrease approximately in inverse proportion to fit-interval length, as illustrated in Fig. C-14 for a fit interval of 10 revs. As a result, the magnitude of this sidelobe in the error spectrum for A for $1/\sqrt{f}$ input noise (upper curve) decreases approximately in proportion to $1/\sqrt{T_{fit}}$ relative to the mainlobe. For example, in Fig. C-14, the largest sidelobe is 0.3 compared to 0.43 in Fig. C-13 (with 5-rev fit interval), which is approximately a factor of $\sqrt{2}$ reduction. Thus, the error contribution of this largest sidelobe is even smaller for longer fit intervals. For a 7-day fit interval, sidelobe magnitude would be reduced by a factor of $\sqrt{7 \times 16/10}$ or 3.3 relative to Fig. C-14 with a 10-rev fit interval. (Type-II resonance sidelobes do not decrease in this fashion as fit interval increases.) This heuristic example suggests that expected $1/\sqrt{f}$ accelerometer noise in the acceleration observable can be modeled reasonably well with one constant for the entire fit interval (e.g., 7 days), at least for purposes of approximate error analysis. This result supports the simulations of Section 6.

When the instrumental model is either a piecewise constant or one constant for the whole fit interval, the relative magnitudes of the mainlobes in a fit filter for a multitone partial are the same as the relative amplitudes of the tones in the partial. For example, when the partial is a sum of 1/rev and 3/rev cosines of equal (unit) amplitude and the instrumental model is a piecewise constant with 1 rev updates, the fit filter is similar to the filter in Fig. C-9 for which no instrumental model is estimated. That is, the magnitudes of the two mainlobes are each nearly 0.5 and the sidelobes drop off relatively rapidly. Thus, for the case of the corrected acceleration observable with a constant instrumental model, the frequency placement and magnitudes of mainlobes in the relevant fit filters are given directly by the frequency placement and amplitudes of tones in the partial and excessive sidelobes are not a substantial concern. This result supports the approach of Section 5 where partial tones are "mapped" directly to fit-filter mainlobes.

C.2.9 Reducing Resonance Sidelobes by Varying Update-Interval Length

The results of Subappendix C.2.4 suggest that a fixed update interval in the range of 1 rev to 2 rev, depending on error characteristics, can provide reasonably good results, but excessive resonance sidelobes in the fit filters indicate a better instrumental model should be feasible. One possibility is to successively vary the length of the update interval over a fit interval (e.g., values such as 1.4, 1.5, 1.6, 1.7 rev, etc. over 13 days). The resulting frequency placement of resonance sidelobes would be different for each update-interval length so that the composite effect in the overall fit filter would be greatly reduced. Fig. C-15 illustrates the change in location of resonance sidelobes relative to the example in Fig. C-3, when the update-interval length is changed from 1 rev to 1.1 rev. Both examples treat a simultaneous fit for a 2/rev-sine amplitude and a piecewise-quadratic instrumental model. In both figures, the fit filter for the sine amplitude has a peak with a magnitude of 1.0 at a frequency of 2/rev, which passes the signal amplitude with unit "gain", as one would expect. Note that corresponding resonance sidelobes in Figs. C-3 (lower fit-filter curve) and C-15 are clearly separated in frequency (e.g., $\Delta f = 0.1$/rev for the corresponding sidelobes near 1/rev). This example suggests one could switch between, say, 5 different update interval lengths (say, 1.4, 1.5, 1.6, 1.7, and 1.8 rev) over a 13-day fit. Relative to the resonance sidelobes found in the case of a single length for all update intervals, this should substantially reduce the magnitude of each resonance sidelobe in the overall fit filter.

152

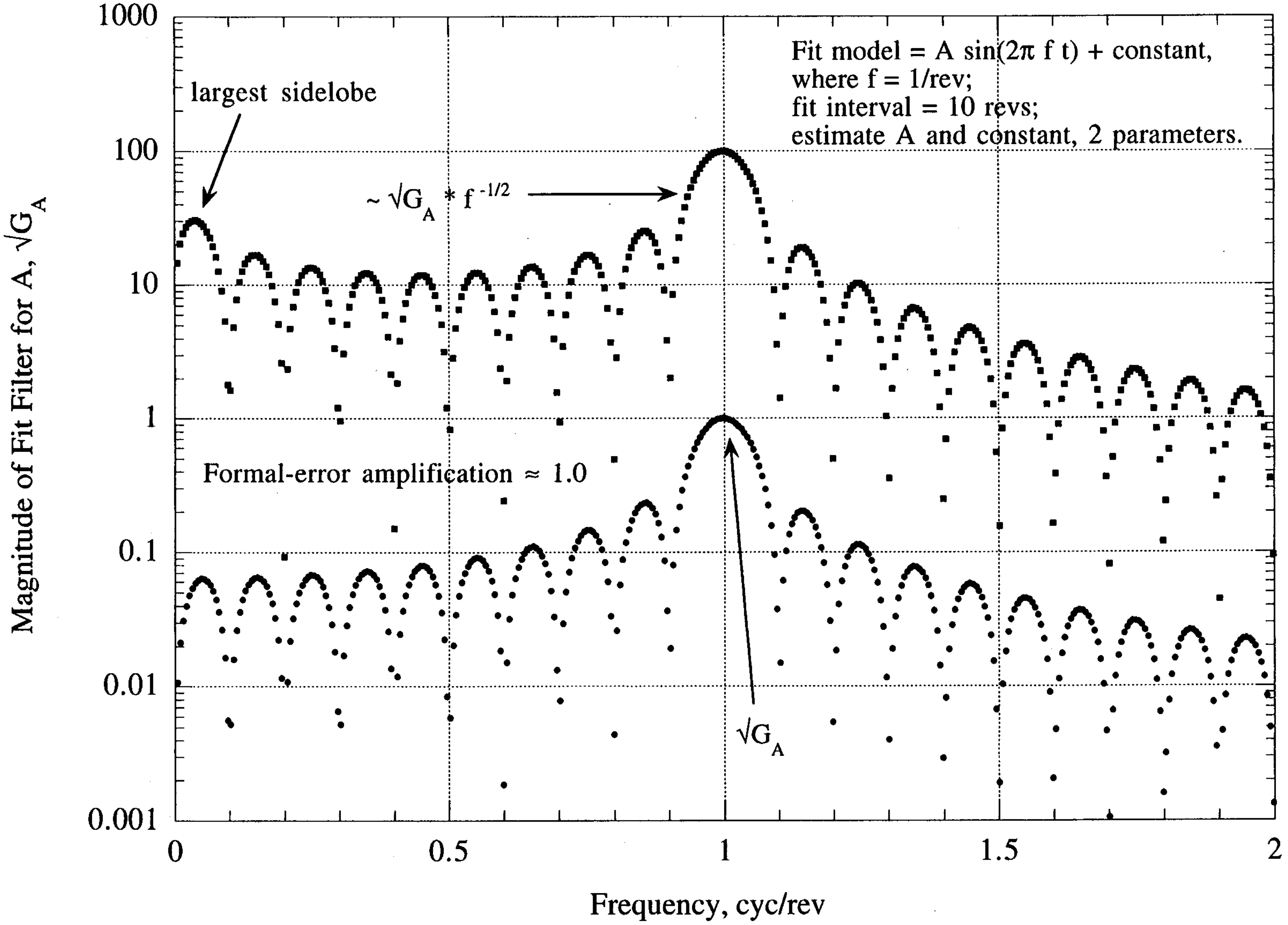

Figure C-14 Fit Filter for the Amplitude of a 1/rev Sine Function when Estimated along with a Constant Over 10 Revolutions

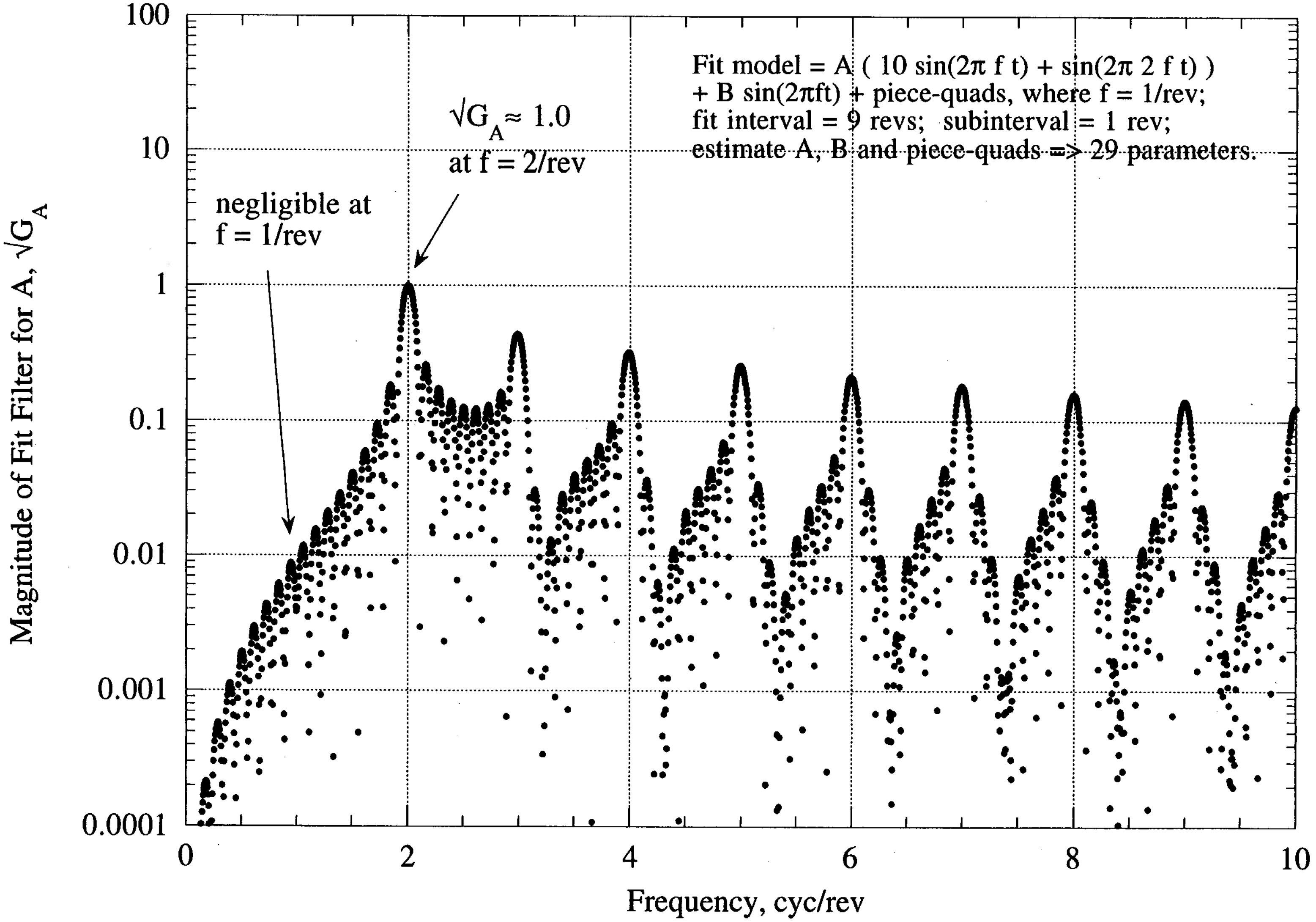

Figure C-15 Fit Filter for the Amplitude of a Two-sine Signal when Estimated along with a 1/rev Sine and a Piecewise Quadratic Function

C.2.10 Two-Tone Signal Plus 1/rev Tone

The following simple example provides a graphic frequency-domain illustration of a somewhat obvious concept with important implications. Suppose that the total signal is comprised of a two-sine signal (1/rev + 2/rev) plus a separate 1/rev-sine signal. In the two-sine signal, let the subamplitude of the 1/rev sine be 10 and the subamplitude of the 2/rev sine be 1 (i.e., $A[10\sin(2\pi f t) + \sin(2\pi 2 f t)]$ where $f = 1/\text{rev}$). In the fit, let the overall amplitude (A) of the two-sine signal, the amplitude (B) of the 1/rev-sine signal and a piecewise-quadratic instrumental function all be simultaneously estimated. Fig. C-16 presents the resulting fit filter for the amplitude A of the two-sine signal. As one would anticipate on the basis of the common $\sin(2\pi f t)$ signature, estimation of the 1/rev sine amplitude (B) "absorbs" the very large 1/rev sine component in the two-sine signal so that the 1/rev peak in the fit filter for A is suppressed. Consequently, the 2/rev peak in the fit filter dominates estimation of A. That is, the frequency-domain fit filter for A takes the form (namely a 2/rev peak with 1.0 magnitude) that would have been generated if only the 2/rev component had been present in the signal, with the form $A\sin(2\pi 2 f t)$. If one considered the partial of the two-sine signal with respect to A outside the context of the fit, one would incorrectly conclude that the large 1/rev sine would dominate the estimation of A. In practice, however, it is inconsequential in the fit filter for A and can be ignored in error analysis when mapping observable errors to errors in amplitude A. Only the 2/rev component and the magnitude of its subamplitude matter in such error mapping.

This simple example suggests that, in a complete fit where orbit parameters, instrumental parameters and gravity coefficients are simultaneously estimated, gravity-coefficient errors are not substantially degraded as a result of orbit-parameter-related signatures in a partial with respect to a gravity coefficient. In complete fits, the partial with respect to a given gravity coefficient contains two categories of components: a) a direct term that results from the explicit gravity-coefficient dependence of the gravity field (e.g., see Eqs. (4.6) and (4.7) in the case of corrected acceleration) and b) an indirect category containing many terms that result from the change of orbital position (e.g., the arguments r, θ, λ) as a result of change in that gravity coefficient. As illustrated in Section 5 for corrected acceleration, the direct term is multitonal with tone frequencies ranging from zero to n/rev and, for purposes of error analysis, is not substantially changed by orbit variations caused by gravity harmonics or by small changes in orbit parameters.

The indirect category includes two subcategories: a first subcategory consisting of 1/rev tones and other effects that can be absorbed by readjustment of orbital parameters (six for each satellite), and a second subcategory consisting of other tones that can be absorbed by adjustment of the gravity coefficient. The first subcategory is a result of the fact that the orbital impact of a given set of orbit parameters changes when any given gravity coefficient changes [Courtney Duncan, private communication, 1997]. For example, a specific set of orbit parameters that would lead to a perfectly circular orbit if only the point-mass (1/r) term were present in the gravity field, can lead to orbits with a 1/rev "elliptical signature" (in addition to the expected 2/rev J2 signature) when J2 is "turned on." Even though this subcategory of indirect sensitivity of orbital position to gravity coefficient is important, the associated orbital effects can be "absorbed" by readjustment of orbit parameters. Thus, the fit filter for J2 is expected to consist primarily of a 2/rev peak.

This example suggests that, even though the indirect orbit-parameter-related terms (first subcategory) can be large in the raw partial for a given gravity coefficient, they can be suppressed in the fit filter for that coefficient as result of concurrent estimation of orbit parameters. That is, concurrent estimation of orbits can cause the appearance of substantial indirect 1/rev signatures and other orbit-parameter-related signatures in the partial with respect to a given gravity coefficient, but such signatures are subsequently suppressed in the fit filter for that coefficient by concurrent estimation of orbit parameters and therefore are not significant in the estimation of that coefficient.

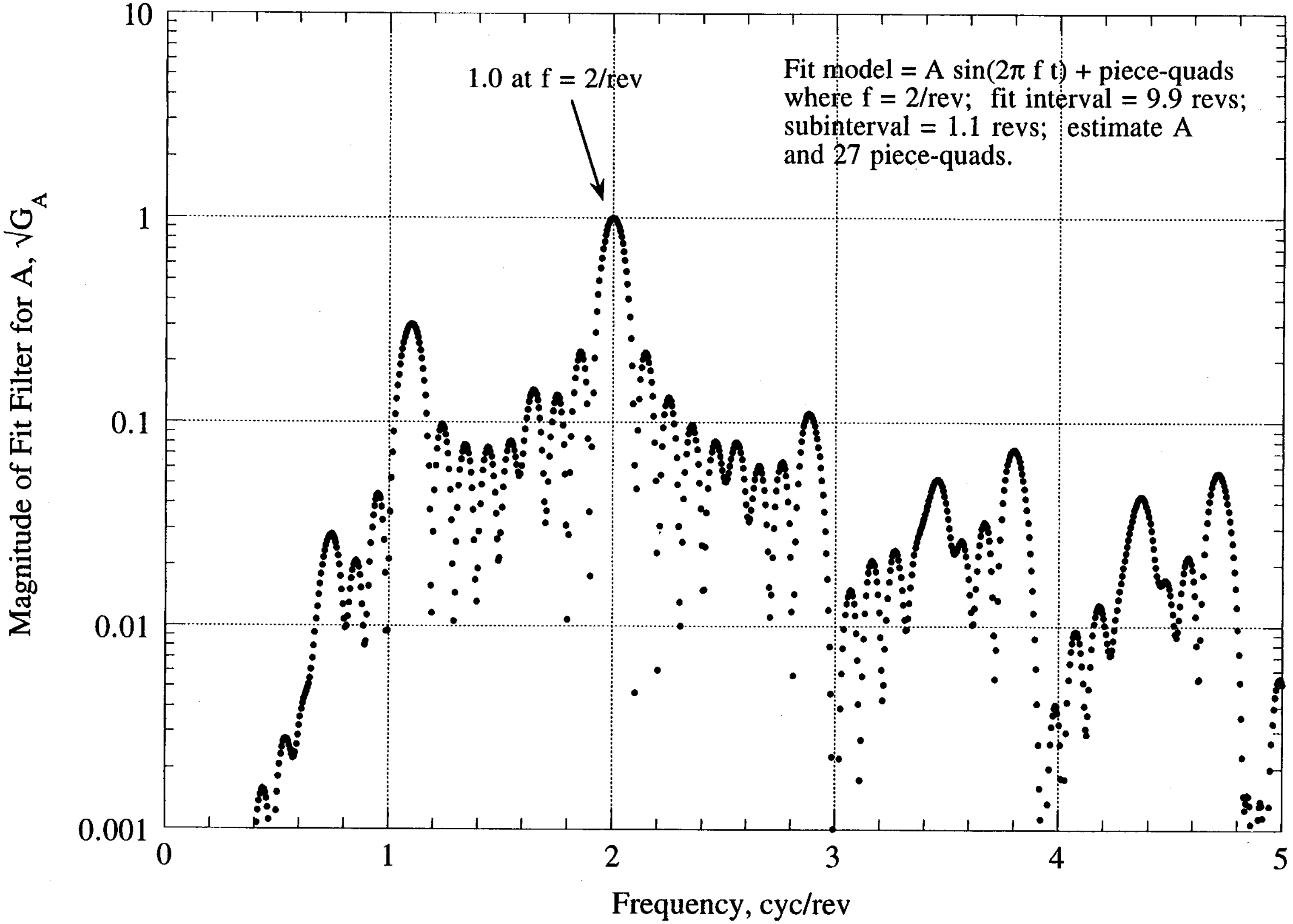

Figure C-16 Fit Filter for the Amplitude of a 2/rev Sine Function when Estimated along with a 1.1-rev-piecewise Quadratic Function

This leaves the other tones in the fit filter to determine the gravity coefficient. It should be noted that a partial for an odd zonal that contains a 1/rev tone as a result of the direct derivative with respect to coefficient, can result in a fit filter that lacks the expected direct-derivative 1/rev peak because of orbit-parameter estimation. In such cases, orbit-parameter estimation can alter the fit-filter tone distribution relative to the direct partial. Such coefficients are the exception since, for example, only one in 200 is an odd zonal, up to degree 100. Further, as indicated in the preceding examples, it should be noted that any tones in the partial that fall at approximately 1/rev in frequency or lower can be heavily suppressed in the fit filter by a 1-rev-update instrumental model.

The approximate approach of Sections 4, 5, and 6 neglects the indirect terms in the gravity-coefficient partials. More analysis is needed to assess these indirect terms, particularly the second subcategory.

C.2.11 Two Highly Correlated Multitone Signals

In a simultaneous fit for many gravity coefficients, relatively high correlations can occur between coefficients with the same (m, j) values (e.g., see Appendix D). These correlations are a result of the fact that the underlying spherical harmonics in a given (m, j) set, and therefore the corresponding partials, can all possess the same tone frequencies, where the tone subamplitudes can be fairly similar for pairs of spherical harmonics. To illustrate an important fit-filter effect of such correlations, a simplified fit has been carried out to simultaneously estimate the overall amplitude for each of two signals, where each signal is a sum of a 1/rev sine and a 3/rev sine, with very closely matched subamplitudes. The fit also simultaneously estimates a piecewise quadratic instrumental function. The correlation in this example is far stronger than correlations found in actual gravity estimation.

The fit filter for the amplitude (B) of one of the signals is shown in Fig. C-17. Note that filter magnitude is now plotted with a linear scale. The most striking feature of this plot is the magnitude of the mainlobes located at 1/rev and 3/rev. These mainlobe magnitudes are to be compared with the mainlobe magnitudes in Fig. C-9, in which the fit estimates only one parameter, the overall amplitude of the sum of a 1/rev cosine and a 3/rev cosine. (The fit filter in Fig. C-9 is very nearly the same as the fit filter for a sum of a 1/rev sine and a 3/rev sine.) The mainlobe magnitudes in Fig. C-17 are approximately 20 times the magnitudes in Fig. C9 as a result of the closely matched subamplitudes given in Fig. C-17. These excessive mainlobes lead to a formal-error amplification of approximately 20 for this estimated amplitude parameter. A similar conclusion can be drawn by comparing Fig. C-17 with Fig. C-10, which is like Fig. C-9 except that a piecewise quadratic is simultaneously estimated along with overall amplitude. The fit filter for the other amplitude (A), which is not shown, would exhibit similar excessive mainlobes.

In Fig. C-17, note the relatively large resonance sidelobes at 2/rev, 4/rev, 5/rev, etc. that are caused by the piecewise quadratic, similar to the effect seen between Figs. C-9 and C-10. The A/B correlation increases the magnitude of these resonance sidelobes along with the mainlobes. For reference, if A and B in Fig. C-17 were estimated without a piecewise quadratic, the correlation between the two parameters would be 0.9986. This is far larger than the largest correlation (≈ 0.9) between gravity coefficients computed for the corrected acceleration observable (see Appendix D). Thus, the mainlobe amplification exhibited in this simple example far exceeds the amplification expected in gravity estimation, at least for the hypothetical acceleration observable.

157

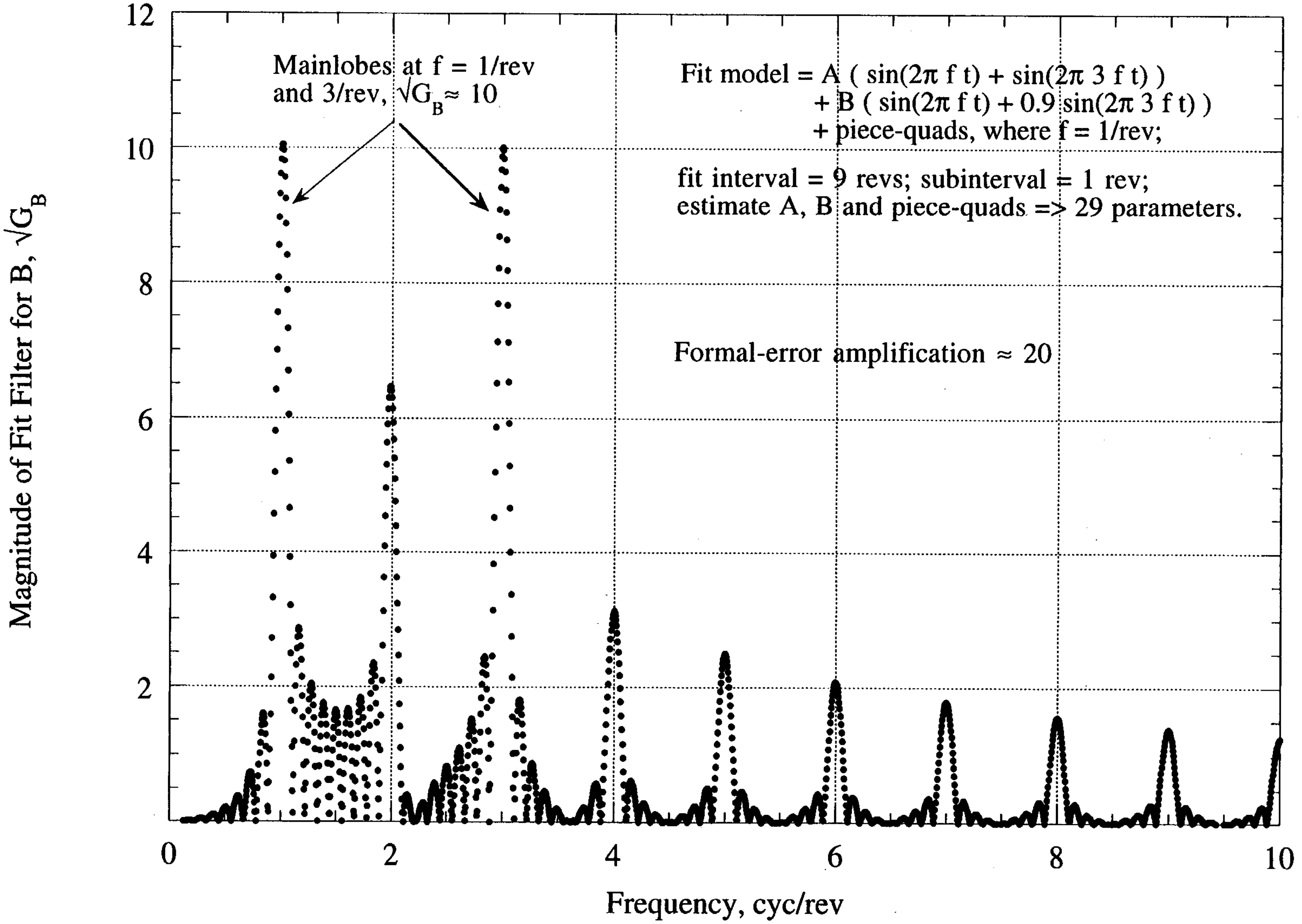

Figure C-17 Fit Filter for the Amplitude of a Two-Tone Signal that is Highly Correlated with Another Signal

Another example of strongly correlated signals is presented in Fig. C-18 in which the two signals are each comprised of four tones (at 1/rev, 3/rev, 5/rev and 7/rev), with the subamplitudes nearly matched between signals. A piecewise-quadratic instrumental function is again estimated along with the overall amplitudes of the two signals. The fit filter for amplitude B is plotted in Fig. C-18, again with a linear scale. This example, which also exhibits exaggerated mainlobes due to the strong correlation between signals, illustrates how ratios of mainlobe magnitudes can change as a result of parameter interactions in the fit. For example, the 1/rev mainlobe is strongly suppressed relative to the other mainlobes (by the piecewise quadratic) and the ratios of the other mainlobes are not the same as the ratios in the partial. Note that resonance sidelobes are also prominent in this example. For a four-tone signal, the single-estimated-parameter mainlobes would each have a magnitude of approximately 0.25 so that formal-error amplification for B in Fig. C-18 is approximately 20.

These examples indicate that strong correlations between gravity coefficients can show up in the fit filter for a given gravity coefficient as large increases in the magnitudes of some of the mainlobes that correspond to the tones in the partial for that coefficient. Large increases in mainlobe magnitudes would lead, of course, to a larger error in the estimated coefficient as a result of magnification of random noise and/or error tones that are located at the frequencies of the amplified mainlobes.

C.3 Appendix Summary

The preceding heuristic examples illustrate the nature of mainlobes and sidelobes found in a frequency-domain fit filter for the amplitude of a signal and suggest that gravity-coefficient errors can be amplified as a result of both excessive sidelobes and excessive mainlobes. Excessive mainlobes in the fit filter for a given gravity coefficient can be caused by strongly correlated gravity coefficients. Such correlations can cause increases in coefficient error, as discussed in more detail in Appendix D on the basis of multiparameter fits to the corrected acceleration observable.

The examples suggest that piecewise-polynomial models can be relatively effective in removing $1/f$ and $1/f^2$ instrumental drifts when properly used. Two types of excessive sidelobes can appear, however: low-frequency (<< 1/rev) sidelobes (Type I) and resonance sidelobes (Type II). The particular heuristic examples analyzed here suggest that a fixed piecewise-interval length in the range of 1 rev to 2 rev for a piecewise polynomial model could suppress the low-frequency excessive sidelobes (Type I) caused by instrumental drifts and provide reasonable performance. However, the analysis suggests that, in some cases, gravity coefficients obtained from fits using a fixed-length-piecewise-quadratic model might be degraded to some extent by resonance sidelobes (Type II), sidelobes that might excessively pass both random noise and/or error tones. A potential, relatively simple approach for substantially reducing resonance sidelobes, if they turn out to be a significant problem, is to vary the length of the piecewise-polynomial update interval over the multiday fit interval (e.g., using lengths of, say, 1.4, 1.5, 1.6, 1.7 revs, etc. in some selected sequence). With this approach, resonance sidelobes generated for each different update-interval length would be located at different frequencies than for other update-length settings and would therefore have much lower magnitudes in the overall multiday fit filter.

A more complete fit-filter analysis is needed to quantify error propagation for actual fits. Such an analysis would be based on actual anticipated observable errors from all sources and on full fits to range and/or range rate, simultaneously extracting orbits and gravity harmonics. Such an analysis is considerably more complicated than the heuristic analysis presented above, but would provide more definitive results. (Due to the large number of gravity coefficients, it would not be reasonable to generate fit filters for all gravity coefficients, but representative examples for each "class" of coefficient might be investigated.)

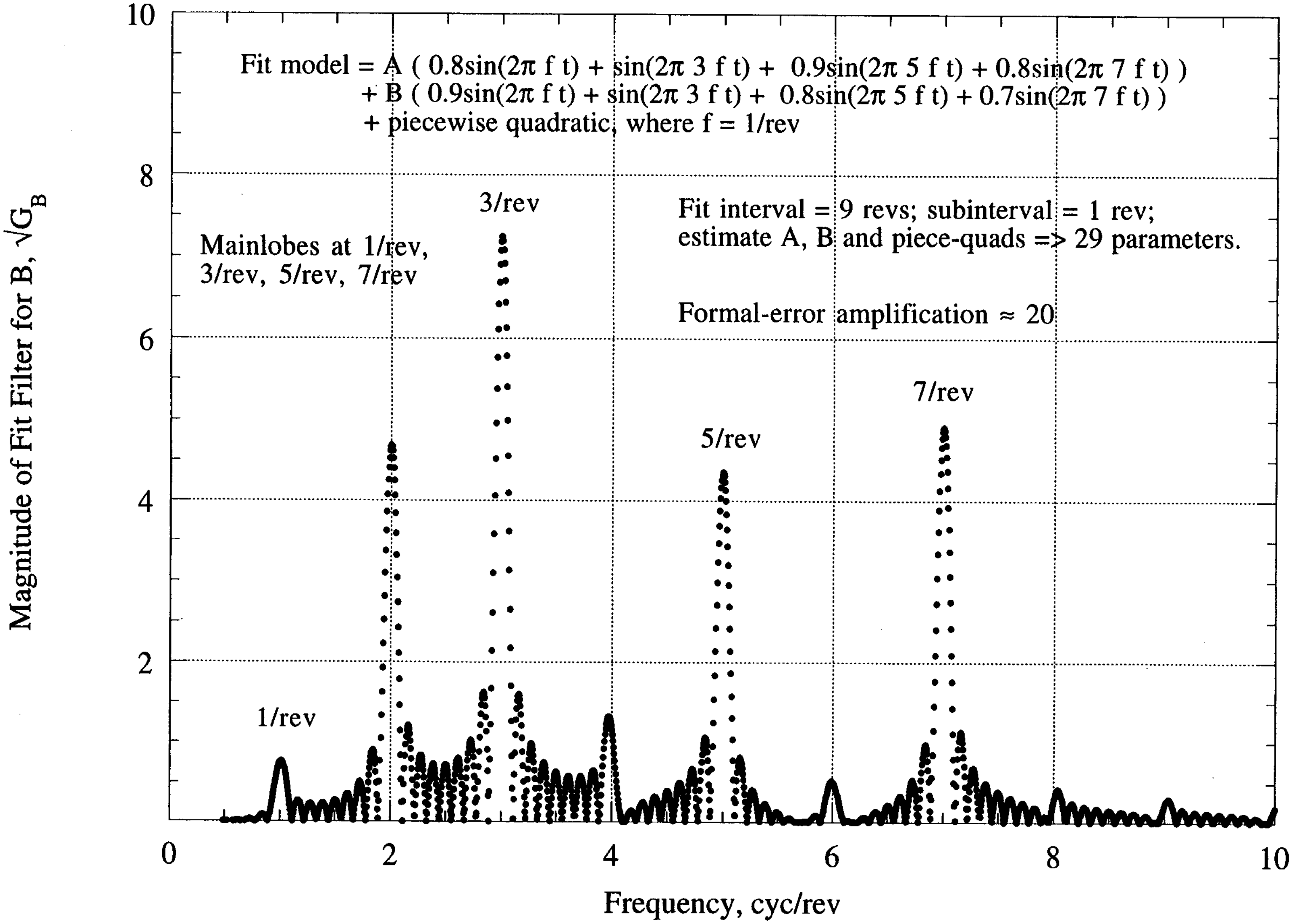

Figure C-18 Fit Filter for the Amplitude of a Four-Tone Signal that is Highly Correlated with Another Signal

Other instrumental models may be better than the piecewise-polynomial approach but it is outside the scope of this report to investigate other approaches. The resonance-sidelobe disadvantage of a piecewise-quadratic model is caused by the quadratic "structure" enforced over each update interval. A better model might be more "stochastic" in nature, without piecewise polynomial structure but with a response "time constant" on the order of 1 rev, so that the fit filters would not exhibit excessive "resonance" sidelobes but would still remove long term drifts.

The partial for a given gravity coefficient in actual gravity estimation can be comprised of many tones, including significant tones at relatively low frequencies (< 1/rev). The preceding examples indicate that such low-frequency tones can be greatly attenuated in the respective fit filter as a result of a piecewise-polynomial instrumental model with, say, 1 rev updates and therefore can be greatly reduced in importance. As a heuristic example, note that the 1/rev tone found in the partial in Fig. 11 is reduced in the fit filter in Fig. C-12 by approximately a factor of ten with respect to its relative magnitude in the partial (i.e., from 0.2/0.2 to 0.026/0.24), as a result of a piecewise quadratic instrumental model with 1 rev updates. The reduction would be even more dramatic if the tone were at lower frequency (e.g., < 0.5/rev). Thus, the lowest frequency tones (e.g., < 1/rev) in multitone partials can have far less significance in affecting the outcome of the fit (i.e., in propagating observable errors) than their magnitudes in the raw partials would suggest. Higher-frequency tones in the partial for a given parameter generally show up as mainlobes in the fit filter for that parameter. Correlations between parameters in a multiparameter fit can cause the mainlobes in the fit filter for a given parameter to be amplified relative to mainlobe magnitudes expected for a single-parameter fit, along with a corresponding increase in the formal error.

In a raw partial for a specific gravity coefficient, computed for a fit that estimates orbit, gravity coefficients and a piecewise-quadratic instrumental model with, say, 1-rev updates, imposing tones at 1/rev and other orbit-parameter-related signatures can have far less significance in the fit filter for that coefficient than relative magnitudes in the raw partial might suggest. The fit-filter example of Subappendix C.2.10 indicates that concurrent estimation of orbit parameters suppresses such orbit-parameter signatures in the fit filter for a gravity coefficient, leaving the effect of the direct partials with respect to gravity coefficient, which supports the approach of Sections 4, 5 and 6 where orbits are assumed to be known. However, more analysis is needed to determine the effect of the other indirect terms in gravity-coefficient partials caused by orbit variation.

With regard to the hypothetical acceleration fits, the examples of Subappendix C.2.8 suggest that the zero-correlation approximation and fit-filter assumptions of Section 5 comprise a reasonable and useful approach.

APPENDIX D

CORRELATIONS BETWEEN ERRORS IN ESTIMATED PARAMETERS

This appendix investigates correlations between errors in parameters estimated with a multiparameter fit to the hypothetical acceleration observable (see Section 4).

D.1 Estimated-Parameter Correlations in a White-Noise Fit

D.1.1 Correlation Analysis

The correlation analysis of this appendix is based on a "white-noise fit" applied to observables (y_i) corrupted only by white-noise errors. White noise is defined as noise that is uncorrelated point to point with uniform variance (i.e., $<\delta y_i \, \delta y_j> = \sigma_y^2 \, \delta_{ij}$). In a white-noise fit (to be referred to here as an "unweighted fit"), the weight matrix for the observables is set equal to a unit matrix and the estimated parameters are computed using the well-known expression given by Eq. (5.1). (A "white-noise fit" is optimal when the observables are corrupted only by white noise. However, because of simplicity, white-noise fits are often applied when observable noise is more complicated.) For white noise on y_i, the predicted variance ("formal error"), σ_k^2, on the k^th^ estimated parameter is given by

$$\sigma_k^2 = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}_{kk} \sigma_y^2 \tag{D.1}$$

The off-diagonal elements of $\left(\mathbf{A}^T\mathbf{A}\right)^{-1}$, when properly normalized by diagonal elements (in the form $M_{kk'}/\sqrt{M_{kk}\,M_{k'k'}}$ where $M_{kk'} \equiv \left(\mathbf{A}^T\mathbf{A}\right)^{-1}_{kk'}$), predict correlations between errors on estimated parameters in the case of white observable noise.

It is emphasized that a rigorous quantitative interpretation of the results of this correlation analysis is possible only in the case of an unweighted fit to observables that are corrupted only by white noise. Since actual observables will be corrupted by random errors that are not white and by important systematic errors with various signatures, actual observable errors will not meet the assumptions behind the derivation. Consequently, the results are to be viewed as suggestive and illustrative, and do not necessarily provide an accurate quantitative assessment of actual fits. As indicated in Appendix C, colored noise can map more favorably or less favorably than white noise for a given gravity coefficient, depending on the shape of the colored-noise spectrum and the frequency placement of lobes (mainlobes and sidelobes) in the fit filter for that coefficient.

One can normalize each partial by its RMS (which can be viewed as a rescaling of parameter units) so that the elements of the "normalized $\mathbf{A}^T\mathbf{A}$" can be written as

$$(\mathbf{A}^T\mathbf{A})_{jk} = \cos(\alpha_{jk}) \tag{D.2}$$

where the "cosine between partials" is given by the normalized dot product,

$$\cos(\alpha_{jk}) \equiv \frac{\dfrac{\partial y}{\partial p_j} \bullet \dfrac{\partial y}{\partial p_k}}{\left|\dfrac{\partial y}{\partial p_j}\right| \; \left|\dfrac{\partial y}{\partial p_k}\right|} \tag{D.3}$$

in which each partials vector is defined by

$$\frac{\partial \mathbf{y}}{\partial p_j} \equiv \left(\frac{\partial y_1}{\partial p_j}, \frac{\partial y_2}{\partial p_j}, \ldots\ldots, \frac{\partial y_{N_o}}{\partial p_j}\right) \tag{D.4}$$

When $\mathbf{A}^T\mathbf{A}$ is normalized in this fashion, the k^th^ diagonal element of its inverse (to be referred to as the normalized $(\mathbf{A}^T\mathbf{A})^{-1}$) represents the multiplicative factor by which the error (variance) of the kth parameter increases, relative to a single-parameter fit, as a result of correlations in the fit, when white noise propagates through an unweighted fit. The square root of the kth diagonal element will be referred to as the formal-error amplification for the kth parameter (as in Appendix C). A large value (i.e., substantially greater than 1.0) for the formal-error amplification for a given parameter indicates the presence of excessive lobes (mainlobes or sidelobes) in the fit filter for that parameter (see Appendix C). To obtain correlations between estimated parameters, the additional normalization described above, $M_{kk'}/\sqrt{M_{kk}\,M_{k'k'}}$, must still be applied.

D.1.2 Correlations in the Hypothetical Fit to Acceleration

This subsection applies the preceding equations to parameters found in the hypothetical approach of Section 4. To provide representative examples, the partial "dot product" in Eq. (D.3) has been computed for a selected set of gravity coefficients, each based on the same 50x50 unweighted fit to corrected acceleration over approximately 7 days, given coplanar polar orbits at 300 km altitude and 200 km satellite range. Acceleration partials are computed as outlined in Section 4, neglecting the "centrifugal" term. Estimated parameters include gravity coefficients and piecewise-constant instrumental parameters, with the orbits taken as known quantities. Orbit period is selected to provide good sampling in longitude over the 7-day fit (i.e., with a commensurability ratio of period to sidereal day of 111/7).

Figs. D-1 through D-7 present the dot products, Eq. (D.3), of the normalized partials for selected reference gravity coefficients (i.e., selected rows in normalized $\mathbf{A}^T\mathbf{A}$). To allow log plots, the magnitude of $\cos\alpha$ is plotted. For the (2, 0, 1) coefficient (J2) in Fig. D-1, the largest off-diagonal element (ODE) is with J4 ($\cos\alpha \approx +0.25$) followed in order by J6 (+0.12), J8 (+0.07), etc. These largest ODEs are all positive.

In Figs. D-2 and D-3 presenting results for J39 and J40, respectively, some ODEs are much larger. For J40 in Fig. D-3, ODEs are large and positive for adjacent even JNs (e.g., $\cos\alpha \approx +0.56$ for J38 and J42, +0.41 for J36 and J44, and +0.33 for J34 and J46). Note that there are quite a few (13) ODEs that are relatively large (≥ 0.1) and their magnitude drops off relatively slowly as a function of degree relative to the reference coefficient. J39 in Fig. D-2 has a similar plot, but with large positive ODEs for adjacent odd JNs. One expects such odd, even results because of the quadrature relationship of even and odd Legendre polynomials as a function of latitude. That is, the partial for a coefficient of odd degree (sine dependence) is essentially orthogonal to the partial for a coefficient of even degree (cosine dependence). One also expects the small values found in this plot for coefficients of nonzero order since the sinusoidal variation with longitude for those coefficients is orthogonal to no variation with longitude for order zero, given good sampling on the unit sphere. More generally, harmonics of different order m and/or different j (i.e., $\cos m\lambda$ or $\sin m\lambda$) are also essentially orthogonal.

164

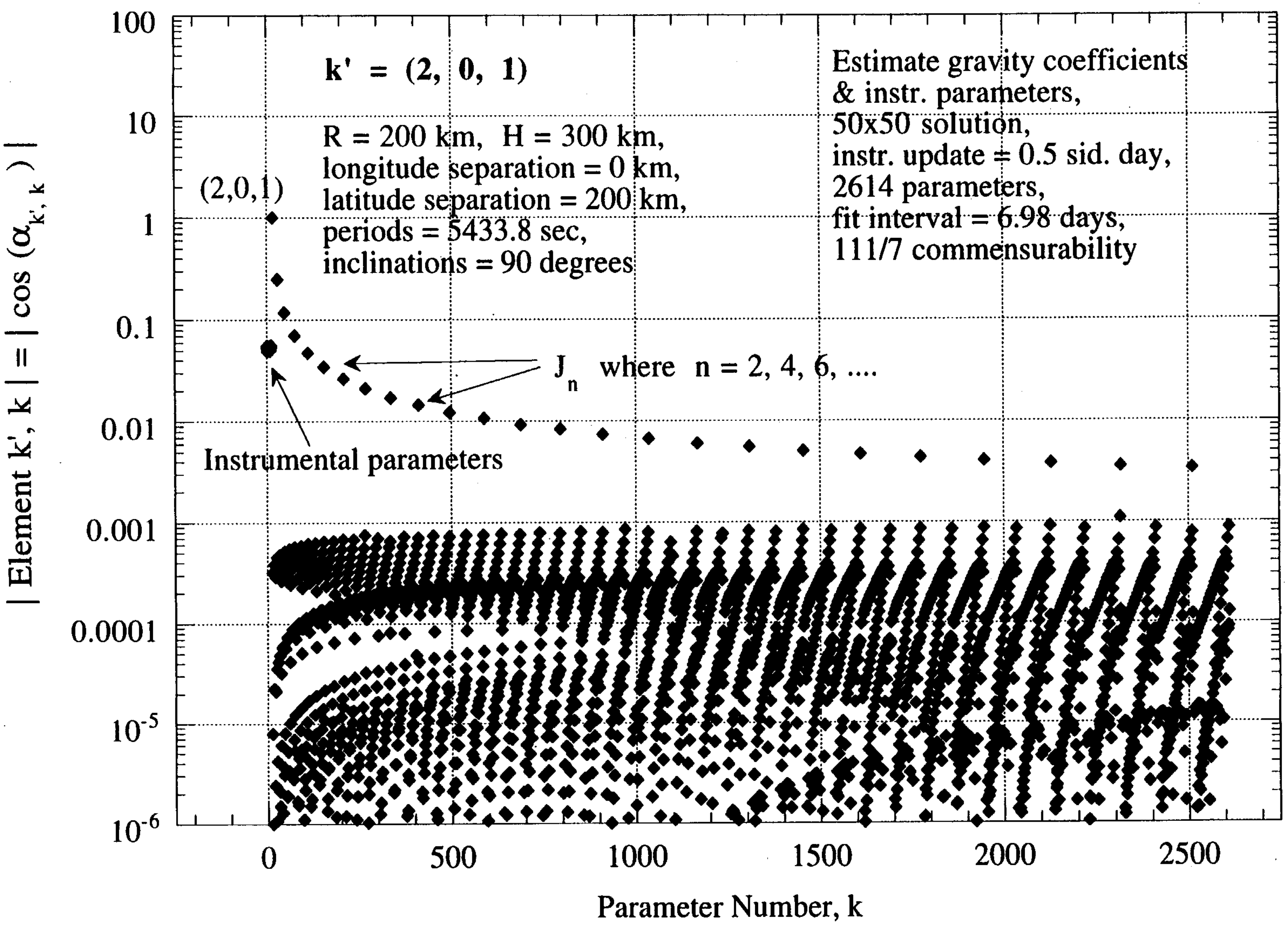

Figure D-1 Elements in Normalized A^TA, Row with k' = (2, 0, 1) = J_2, for the Corrected Acceleration Observable

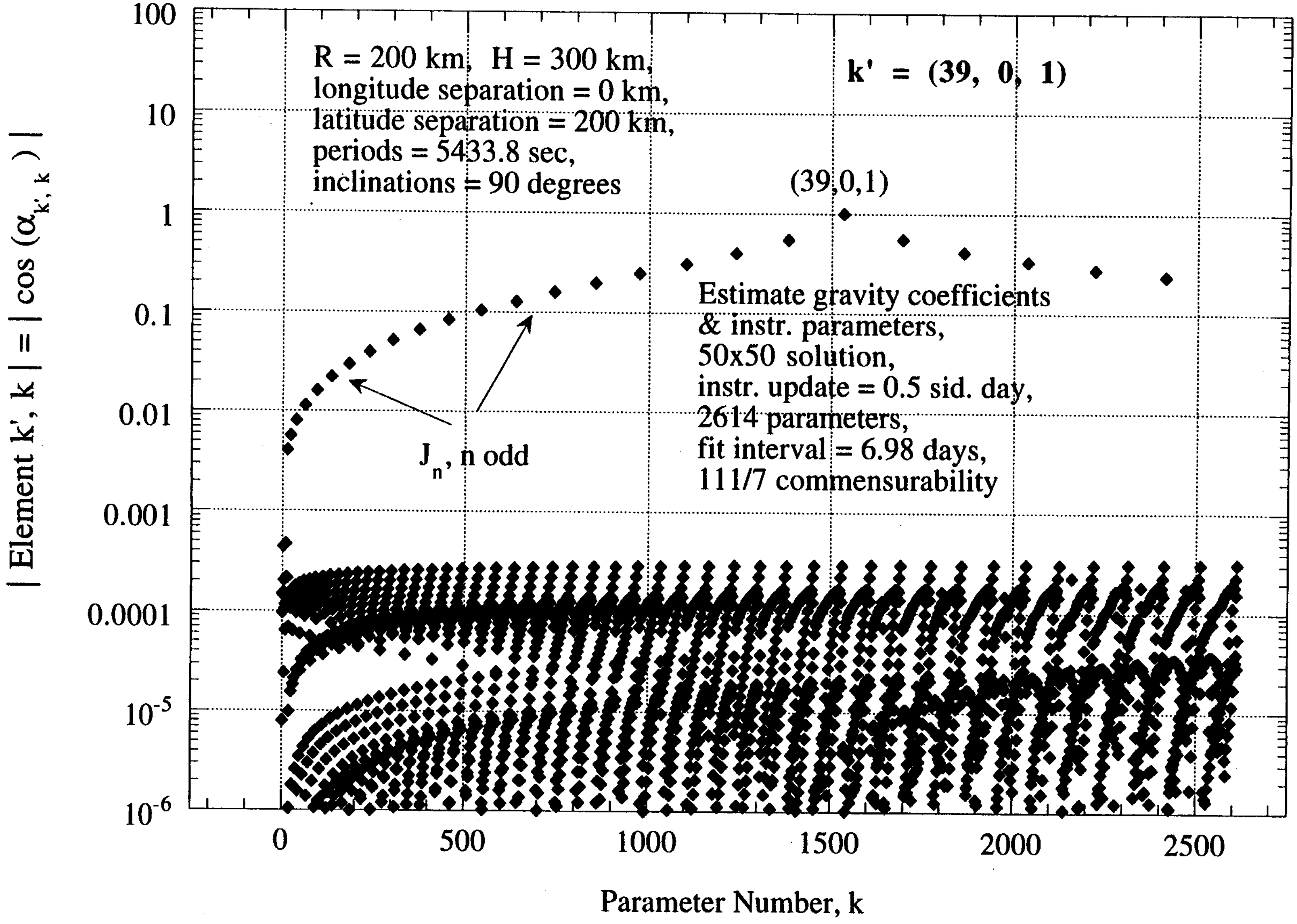

Figure D-2 Elements in Normalized A^TA, Row with k' = (39, 0, 1) = J_{39}, for the Corrected Acceleration Observable

160

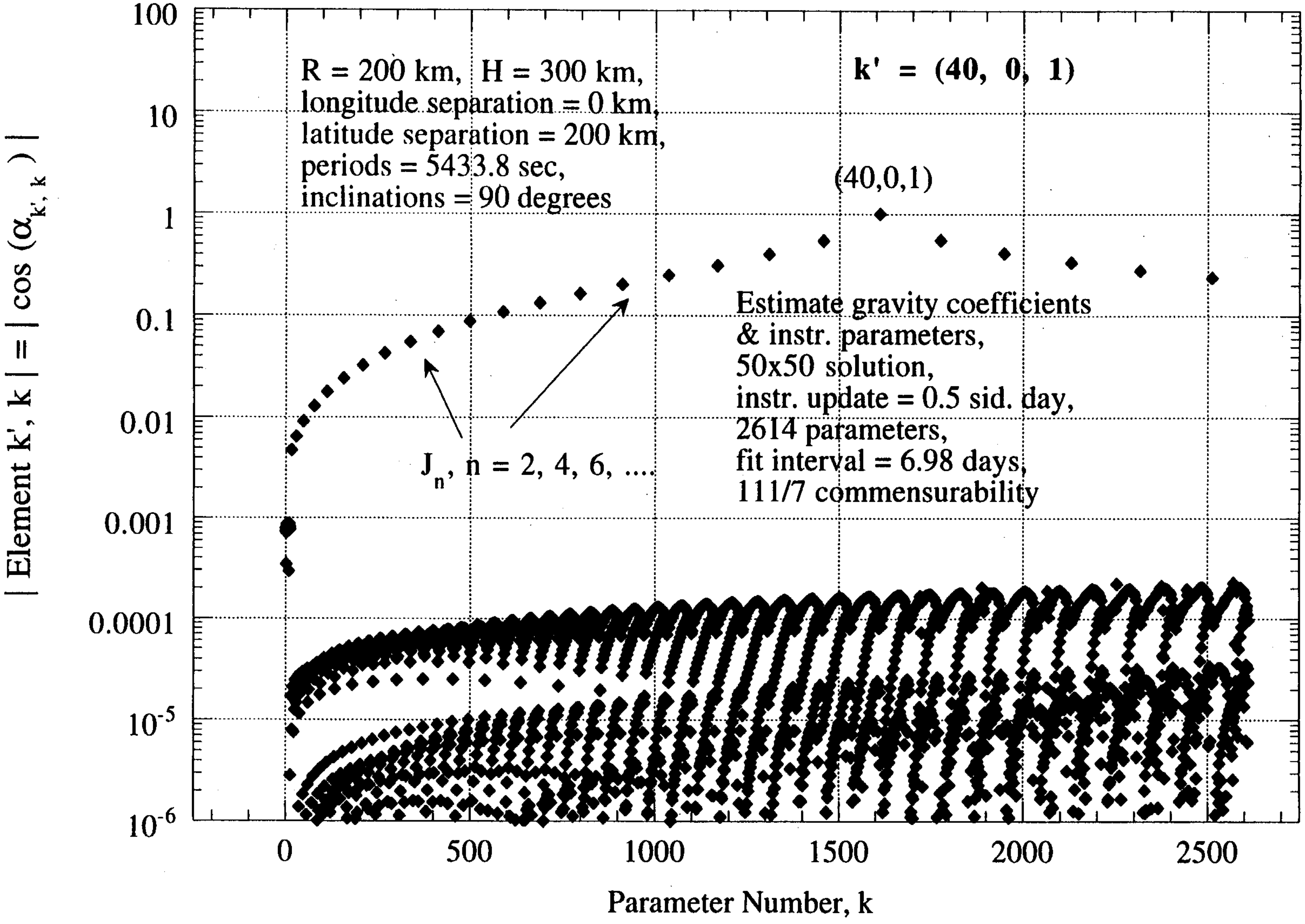

Figure D-3 Elements in Normalized A^TA, Row with k' = (40, 0, 1) = J_{40}, for the Corrected Acceleration Observable

Fig. D-4 presents results for (40, 1, 2), which is similar to J40 in terms of a number of large ODEs. In this case, however, the large values correspond to (n, 1, 2) coefficients, for n even and adjacent to 40, as one would expect for the reasons expressed above.

ODEs are smaller for the (40 , 20, 2) case in Fig. D-5: the projections of (40, 20, 2) on (38, 20, 2) and on (42, 20, 2) are each approximately $\cos\alpha \approx -0.35$, while the projections on (36, 20, 2) and (44, 20, 2) are each approximately –0.13.

Figs. D-6 and D-7 present results for (40, 39, 2) and (40, 40, 2), respectively. For these harmonics, the largest ODEs correspond to other harmonics with the same (m, j) values, as one would expect, but with higher, even degree. (Lower-degree, even harmonics do not reach order 39 or 40.) For example, for (40, 40, 2), the ODEs are –0.52 with respect to (42, 40, 2) and –0.08 with respect to (44, 40, 2).

One distinguishing feature of these dot-product plots is that only a few ODEs out of 2600 have relatively large values (almost all are less than 0.001 and most less than 0.0001). These small values are a consequence of the near orthogonality of partials with different (m, j) values and/or different evenness/oddness in degree value, as discussed above. This orthogonality suggests that the $\mathbf{A}^T\mathbf{A}$ matrix should be partitioned according to (m, j) and evenness/oddness of degree. As an approximation, each partition can then be analyzed separately under the assumption that ODEs "connecting" partitions are negligibly small. Since the size of each partition is relatively small (< 50), $\mathbf{A}^T\mathbf{A}$ for each partition can be easily inverted to determine formal errors and correlations between parameters for the parameters of that partition. This approach provides the exact $(\mathbf{A}^T\mathbf{A})^{-1}$ that would be obtained if only the partition-specific parameters were actually estimated in the fit. The approach can underestimate formal-error amplifications if interpartition ODEs are too large.

Results for formal-error amplification, which are computed as square root of the diagonal elements of normalized $(\mathbf{A}^T\mathbf{A})^{-1}$, are presented in Fig. D-8 for the gravity coefficients found in three selected partitions, namely the allowed even degree values for the (m, j) cases of (0, 1), (20, 2) and (40, 2). For each partition, instrumental parameters have also been included as estimated parameters to make the analysis more complete for the "estimated" coefficients. For the even-degree values (zonals) associated with the (n, m, j) = (n, 0, 1) case, noise is magnified by a factor of only 1.4 or less as a result of fit correlations. For the (n, 20, 2) case, noise is magnified by 1.2 to 2.9, depending on degree value. For (n, 40, 2) case, noise is magnified by 1.3 to 2. Although formal-error amplification can be relatively large for some coefficients, sectorial coefficients (e.g., (20, 20, 2) and (40, 40, 2) in these examples) are magnified by a factor of only 1.2 to 1.3.

Since sectorials dominate the per-degree geoid errors, these results for formal-error amplification in an unweighted fit to corrected acceleration suggest that fit correlations do not substantially degrade per-degree coefficient errors relative to single-parameter fit estimates and that the zero-correlation approximation to error analysis applied in this report should provide reasonable estimates for per-degree geoid errors. However, the relatively large formal-error amplification for some coefficients (e.g., x2.85 for the (32, 20, 2) coefficient in Fig. D-8) suggests the presence of excessive lobes in the fit filters for those coefficients (e.g., see Subappendix C.2.11 for a simple example of excessive mainlobes). Excessive lobes (mainlobe or sidelobe) would make those coefficients particularly susceptible to observable noise and/or observable error tones possessing frequencies that fall on the lobes.

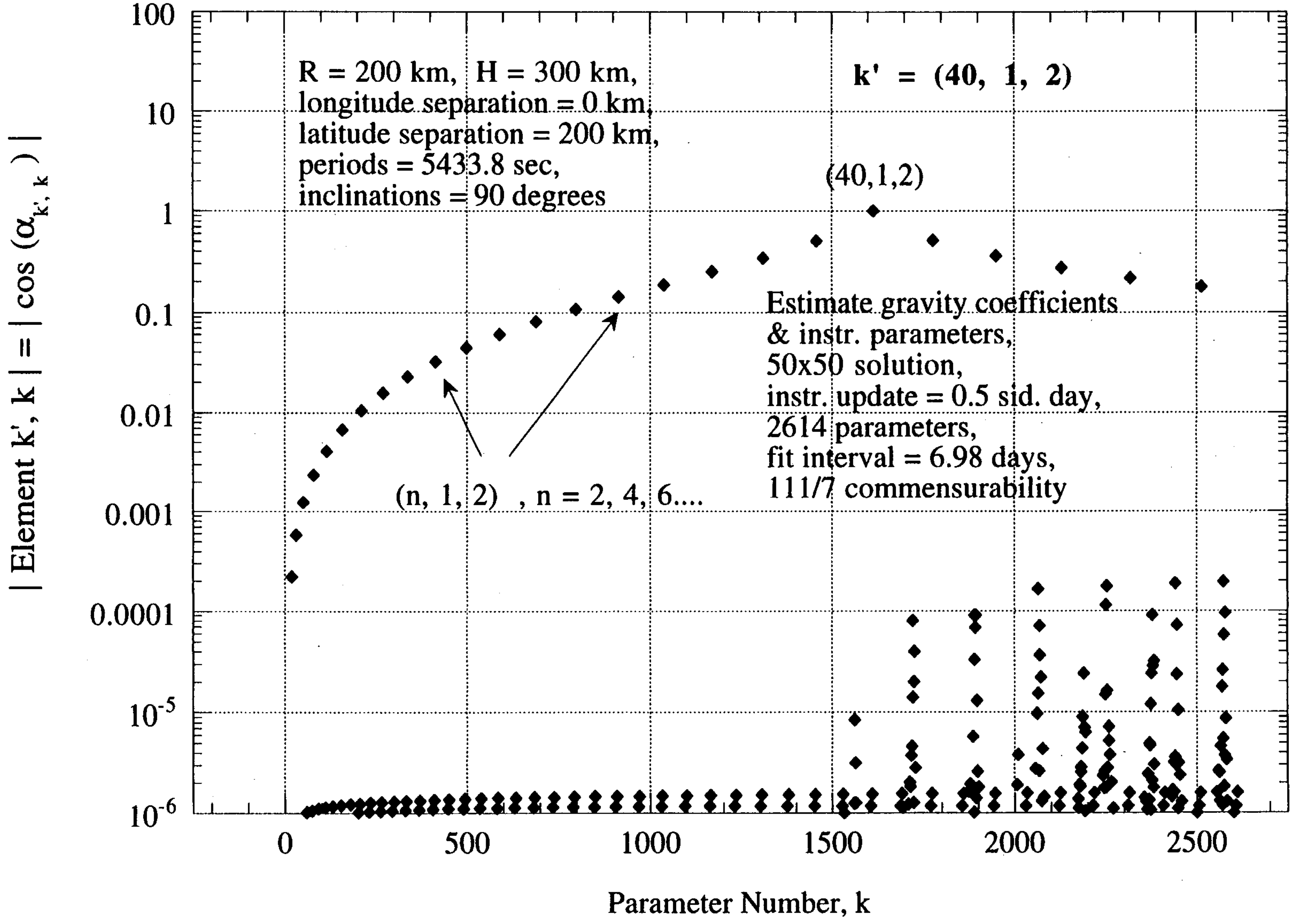

Figure D-4 Elements in Normalized A^TA, Row with k' = (40, 1, 2), for the Corrected Acceleration Observable

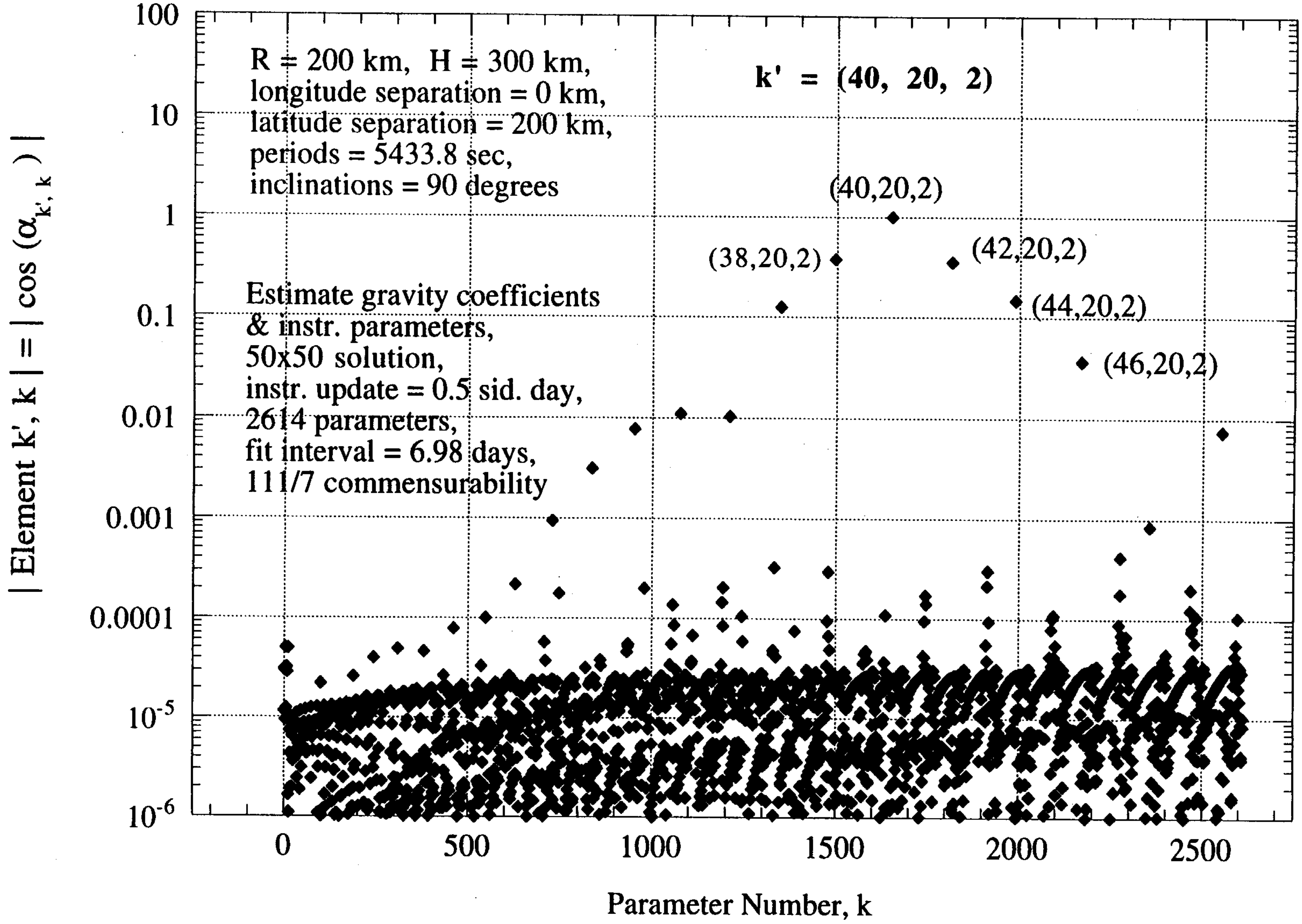

Figure D-5 Elements in Normalized A^TA, Row with k' = (40, 20, 2), for the Corrected Acceleration Observable

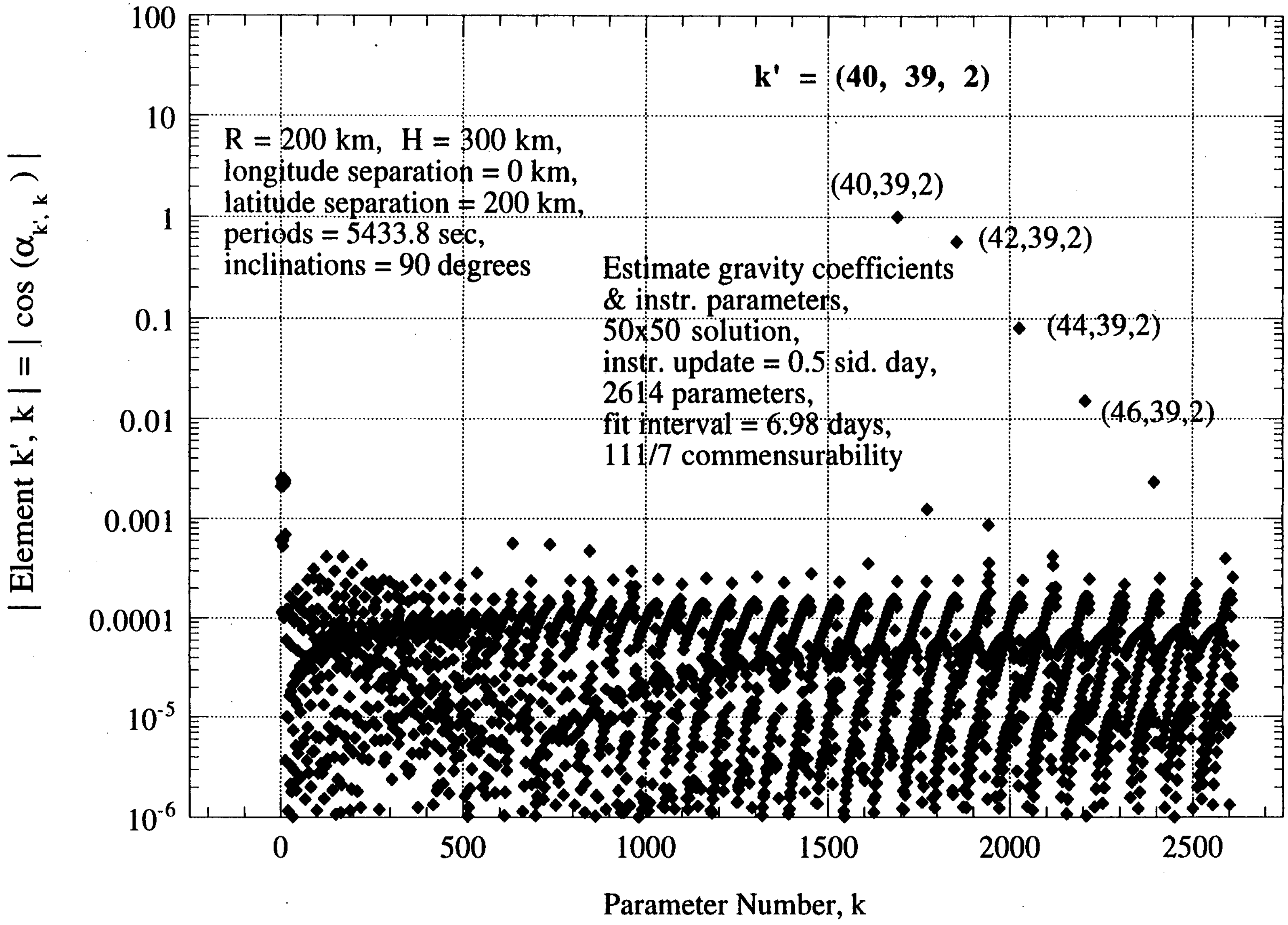

Figure D-6 Elements in Normalized A^TA, Row with k' = (40, 39, 2), for the Corrected Acceleration Observable

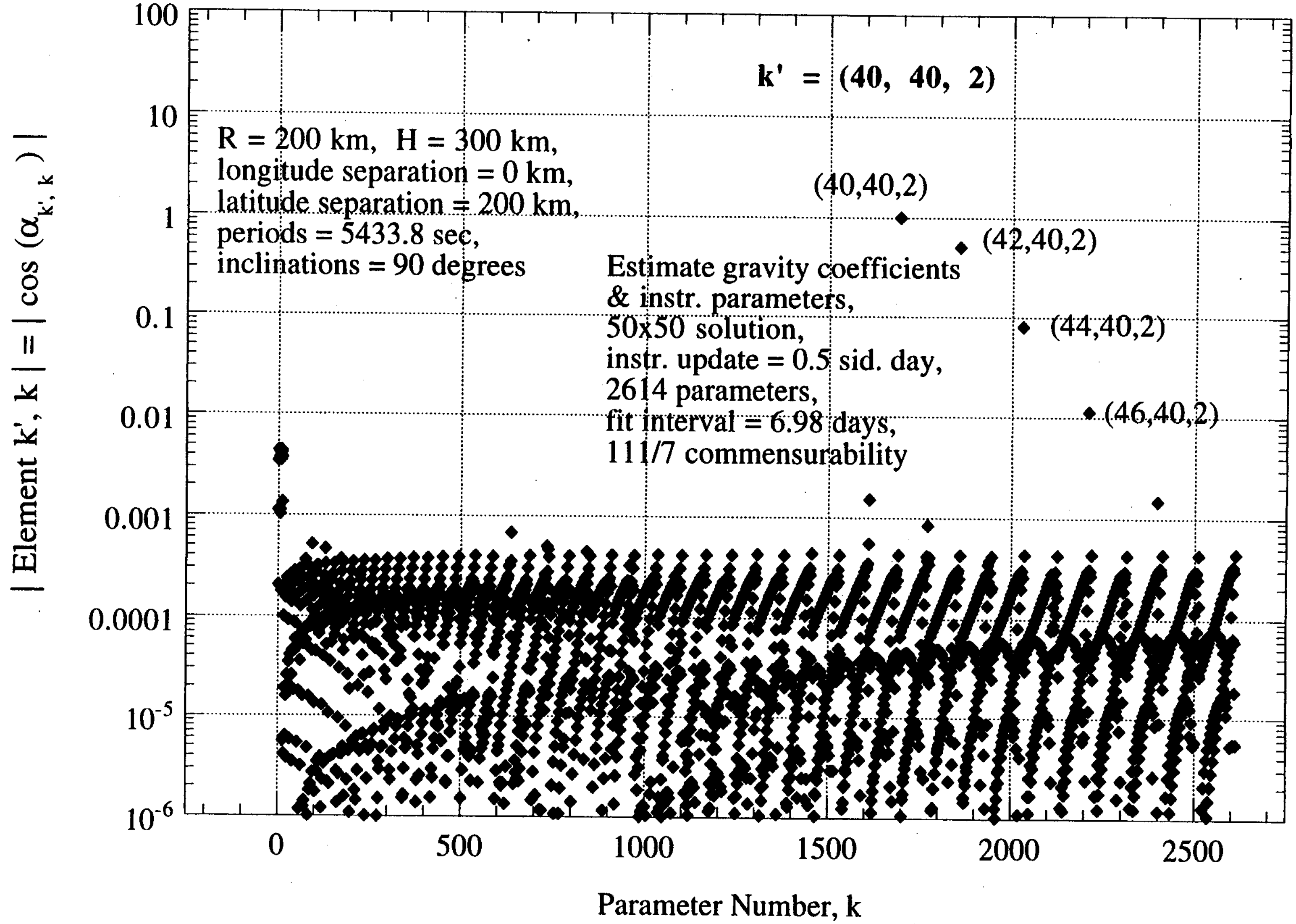

Figure D-7 Elements in Normalized A^TA, Row with k' = (40, 40, 2), for the Corrected Acceleration Observable

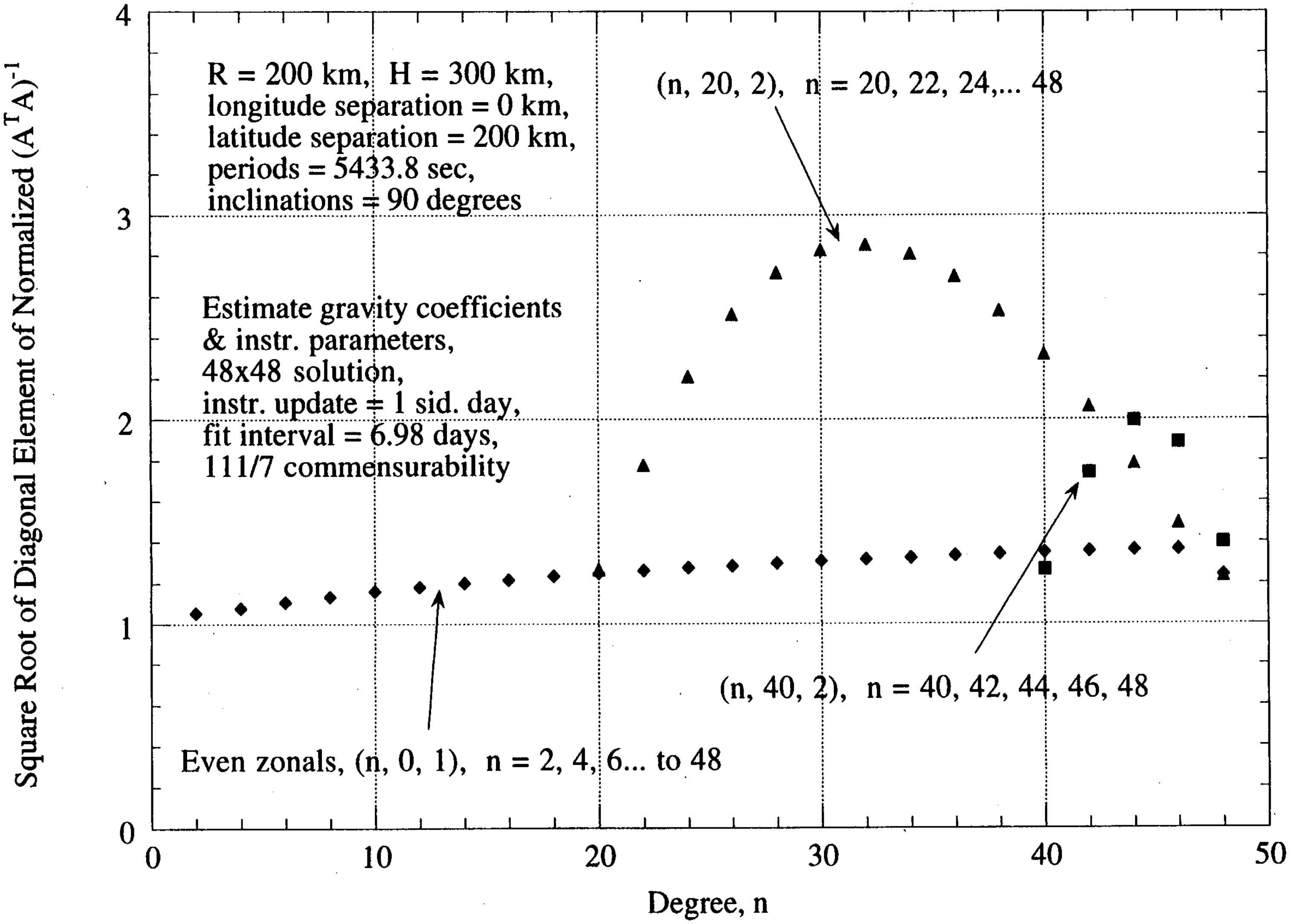

Figure D-8 Approximate Formal-Error Amplification for Selected Gravity Coefficients Estimated with Corrected Acceleration Observable

Correlations of errors between estimated gravity coefficients can also be extracted for each partition from these normalized $(\mathbf{A}^T\mathbf{A})^{-1}$ matrices on the basis of $M_{kk'}/\sqrt{M_{kk}\,M_{k'k'}}$. For the (40, 0, 1) coefficient (J40), the correlations with the three adjacent lower-even-JN coefficients, (38, 0, 1), (36, 0, 1) and (34, 0, 1), are –0.33, –0.064, and –0.027, respectively, with similar values for the three higher adjacent even JNs. Thus, for J40 with a relatively small formal-error amplification of x1.35, the largest interparameter correlations are relatively small, as one would anticipate. For the (32, 20, 2) coefficient, on the other hand, the correlations with the five adjacent higher-even-degree coefficients, (34, 20, 2), (36, 20, 2), (38, 20, 2), (40, 20, 2), and (42, 20, 2), are all positive at +0.89, +0.80, +0.70, +0.61, and +0.51, respectively, with slightly lower values for the five corresponding lower adjacent (n, 20, 2) coefficients. Thus, for the (32, 20, 2) coefficient with a large formal-error amplification (x2.85), some interparameter correlations are quite large, as one would anticipate.

A concern with the unweighted fits is the fact that a polar orbit results in excessive sampling per unit area on the unit sphere at the poles compared with other regions (see next subappendix). Such sampling nonuniformity can potentially allow observable errors that are large at the poles to be strongly mapped to the JN coefficients. This concern is heightened when the behavior of JN polynomials at the poles is considered. As shown in Appendix E, the JN Legendre polynomials for high degree reach very large values at the poles relative to other latitudes and therefore lead to disproportionately large sensitivity partials at the poles. In unweighted fits, these large partial values coupled with disproportionately high sampling density can allow observable errors exhibiting polar extremes to map into unnecessarily large errors in the JN coefficients. Thus, correlations, partial behavior, and polar oversampling suggest consideration of downweighting the polar regions in the fit.

D.2 Estimated-Parameter Correlations in a Fit with cos(lat) Weighting

This subappendix analyzes the option of downweighting polar regions with cos(lat) weighting in a fit to corrected acceleration to determine gravity coefficients. This analysis is presented not to promote cos(lat) weighting but to illustrate the concept and to consider the resulting advantages and disadvantages. Whether it turns out to be advisable to downweight the polar regions with this function or some other function will depend on the ultimate outcome of tradeoffs between different types of errors.

D.2.1 Correlation Analysis

In the standard orthogonality/normalization integral for Legendre polynomials (LPs), a cos(lat) factor enters the integrand to downweight the polar regions. Because of the cos(lat) factor, the aforementioned, disproportionately large values for JN polynomials at the poles are suppressed and therefore have much less influence than they would in a uniform-weight integral. Further, the cos(lat) factor is essential for orthogonality. This aspect of LPs suggests that a similar weighting might be applied in gravity fits to intersatellite observables to down-weight polar regions and to more completely "separate" parameters.

Consideration of sample points per unit area on the unit sphere also suggests cos(lat) weighting might be advantageous. With standard unweighted fits, the polar regions are greatly oversampled relative to other areas when all sample points in the fit are projected onto the unit sphere. For the case of a 90°-inclination orbit and a 13-day fit, approximately 200 points fall at latitude ≈ $\pi/2$ while only one point is found at each equatorial crossing point, assuming a uniform longitude distribution (i.e., a "13-day repeat cycle"). Thus, at $\pi/2$, the number of sample points is approximately two orders of magnitude greater than the number of points at a given equatorial point. When such extreme polar oversampling is combined with the aforementioned disproportionately large values for some partials (e.g., JN partials) for the polar areas, errors localized in the polar regions might severely corrupt some estimated parameters unnecessarily (mainly the zonals and near zonals). Based on the LP analogy, cos(lat) weighting of the observables might improve the fit with regard to both off-diagonal elements and polar oversampling.

The standard weighted least-squares solution corresponding to Eq. (5.1) has the form

$$\mathbf{p} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1} \mathbf{A}^T\mathbf{W}\mathbf{y} \tag{D.5}$$

where **W** is a weighting matrix. In order to end up introducing the promising cos(lat) weighting in the partial dot product, let

$$W_{ij} = \cos(\theta_i)\,\delta_{ij} \tag{D.6}$$

where θ_i is latitude of the i^{th} observation. For purposes of correlation analysis, one can reformulate Eq. (D.5) as an unweighted fit to a new (weighted) observable defined by

$$\tilde{y}_i = \sqrt{\cos(\theta_i)}\; y_i \tag{D.7}$$

so that Eq. (D.5) becomes

$$\mathbf{p} = \left(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}}\right)^{-1} \tilde{\mathbf{A}}^T\tilde{\mathbf{y}} \tag{D.8}$$

where

$$\tilde{A}_{ij} \equiv \frac{\partial \tilde{y}_i}{\partial p_j} \equiv \sqrt{\cos(\theta_i)}\,\frac{\partial y_i}{\partial p_j} \tag{D.9}$$

Note that the form of Eq. (D.8) is the same as Eq. (5.1). Thus, based on these redefined observables, the derivation of Subsection D.1.1 can be directly reapplied, except that now the "angle between partials" is defined by

$$\cos(\alpha_{jk}) \equiv \frac{\dfrac{\partial \tilde{\mathbf{y}}}{\partial p_j} \bullet \dfrac{\partial \tilde{\mathbf{y}}}{\partial p_k}}{\left|\dfrac{\partial \tilde{\mathbf{y}}}{\partial p_j}\right| \left|\dfrac{\partial \tilde{\mathbf{y}}}{\partial p_k}\right|} \tag{D.10}$$

where the new partials vector is given by

$$\frac{\partial \tilde{\mathbf{y}}}{\partial p_j} \equiv \left(\frac{\partial y_1}{\partial p_j}\sqrt{\cos(\theta_1)},\; \frac{\partial y_2}{\partial p_j}\sqrt{\cos(\theta_2)},\; \ldots\ldots,\; \frac{\partial y_N}{\partial p_j}\sqrt{\cos(\theta_N)} \right) \tag{D.11}$$

Note that the dot product in Eq. (D.3) has the form

$$\frac{\partial \tilde{\mathbf{y}}}{\partial p_j} \bullet \frac{\partial \tilde{\mathbf{y}}}{\partial p_k} = \sum_i \frac{\partial y_i}{\partial p_j}\frac{\partial y_i}{\partial p_k}\cos(\theta_i) \tag{D.12}$$

174

which now includes the desired cos(lat) weighting and resembles the weighting in the standard Legendre integral. That is, if the $\partial y_i/\partial p_k$ were spherical harmonics rather than observable derivatives, the sum would be very close to zero due to harmonic orthogonality, given adequate sampling over the unit sphere. As suggested in Appendix E, orthogonality in this weighted sum is nearly satisfied by the second derivatives of the Legendre polynomials and therefore by the partials of the acceleration observable with respect to zonal coefficients (see Eq. (5.16)).

D.2.2 Correlations in the Hypothetical Fit to Acceleration

The preceding analysis shows that the earlier, uniform-weight equations for correlations can be applied to the weighted fit but with the effective partials redefined according to Eq. (D.11). In a rigorous interpretation of this approach, however, the resulting $(\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}$ matrix (i.e., $(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}$) predicts hypothetical errors and correlations for parameter estimates that would be produced when the correlation matrix for input noise on $\tilde{y}_i$ is white. This would mean the correlation matrix for the noise on y_i is hypothetically equal to $\mathbf{W}^{-1}$ (i.e., uncorrelated point-to-point and larger at the poles in proportion to $1/\sqrt{\cos(\text{lat})}$). Such $\mathbf{W}^{-1}$ noise will not occur in actual intersatellite observables. (This is reminiscent of the situation for the $(\mathbf{A}^T\mathbf{A})^{-1}$ approach, which requires unlikely white observable noise (on y_i) for rigorous quantitative interpretation.) Thus, since actual observable errors will not satisfy the assumptions, the analysis is again suggestive and not strictly quantitative.

When the plot for J2 in Fig. D-1 is recomputed using cos(lat) weighting (computed as average cos(lat) for the two satellites, $[\cos(\theta_a) + \cos(\theta_b)]/2$), one obtains the ODEs shown in Fig. D-9. For this JN harmonic, weighting has little effect. For J40, however, shown in Fig. D-10, the ODEs drop dramatically. For example, the ODE of J40 with J38 or J42 drops dramatically from approximately 0.55 to 0.02, or by approximately a factor of 30. This result suggests that cos(lat) weighting can reduce the degradation due to interparameter correlations to very low levels for the higher-degree JN harmonics. The small ODEs between zonals with cos(lat) weighting can be explained by the form of the partials in Eq. (5.16) and the near orthogonality of weighted second derivatives of Legendre polynomials, as discussed in Appendix E.

Fig. D-11 presents results for a weighted fit for the (40, 40, 2) harmonic. Comparison with Fig. D-7 indicates cos(lat) weighting has little effect for this harmonic. Since sectorials ($m = n$) have virtually no strength at the polar latitudes (see Appendix E), it is expected that a cos(lat) weighting would have little effect on ODEs or correlations for this harmonic.

These results indicate that cos(lat) weighting could greatly reduce the susceptibility of zonal and near-zonal coefficients to possible observable errors with polar extremes and can very effectively "separate" the partials for zonals and near zonals. As a result, interparameter fit correlations would cause only minor increases in JN-coefficient errors relative to a single-parameter fit. Since the "normal" partials for zonal coefficients are far larger than partials for the sectorial coefficients (assuming nearly zero-longitude separation between satellites, as illustrated in Fig. 4-6), the possible increase in random errors in zonals caused by down-weighting of polar data might not be as serious in an overall sense as it might first appear. For example, as indicated by Fig. E-13 and Eq. (5.16) for the acceleration observable, the RMS partial for the degree-50 zonal is approximately 1.7 times smaller with $\sqrt{\cos(\text{lat})}$ weighting than with uniform weighting. Since the "normal" zonal partials near degree 50 (see Fig. 4-6) for the corrected acceleration observable are approximately 50 times larger than the corresponding sectorial partials (assuming zero longitude separation between satellites), zonal partials with $\sqrt{\cos(\text{lat})}$ weighting would still be on the order of 30 times larger than the sectorial partials. This suggests that random errors for zonals would still be far smaller than random errors for sectorials (on the order of a factor of 30) even when cos(lat) weighting is applied. This comparison does not take into account the additional improvement ($\approx$ x1.2 to 1.3) in formal-error amplification gained by a weighted fit relative to an unweighted fit. This analysis suggests that the price paid for using cos(lat) weighting would be minor and the advantages might be substantial.

D.3 Appendix Summary

For the hypothetical acceleration fits, large correlations between gravity coefficients can be confined to relatively small partitions in parameter space if A^TA is partitioned according to (m, j) and evenness/oddness of degree. According to the partitioned A^TA approximation, interparameter correlations can be as large as approximately 0.9 (e.g., between (32, 20, 2) and (34, 20, 2)) and can cause formal-error amplifications as large as approximately x3 (e.g., for (32, 20, 2)). Interparameter correlations and formal-error amplifications are relatively small for the sectorial coefficients (e.g., the formal-error amplifications for (20, 20, 2) and (40, 40, 2) are approximately x1.25).

The results of this approximate correlation analysis suggest that the zero-correlation approximation of Section 5, in which errors for each gravity coefficient are computed separately from the others in a single-estimated-parameter approximation, is accurate enough for the crude error analysis based on hypothetical acceleration fits. Further, these results indicate that, at least for the hypothetical acceleration fit, cos(lat) weighting can "separate the partials" for high-degree, low-order coefficients without causing unacceptable increases in coefficient random errors. Further, cos(lat) weighting would compensate for polar oversampling and thereby reduce possible disproportionate mapping of observable errors that exhibit polar extremes. More work is required to compare error tradeoffs between weighting options, particularly in full fits to range and range rate. Extension to these other observable types with simultaneous estimation of orbits and other parameters may significantly change conclusions concerning error tradeoffs and correlations.

According to current GRACE mission plans, actual satellite orbits will not have 90° inclination, such as those analyzed above, but approximately 87° inclinations. By leaving a circle at each pole unsampled, such an inclination would, in effect, downweight the most extreme polar latitudes and change the correlations for JN coefficients. Even for such orbits, however, cos(lat) weighting can still reduce the disproportionate magnitude of JN polynomials at high latitudes. Further analysis is needed to determine the advantages of cos(lat) weighting, if any, for such inclinations.

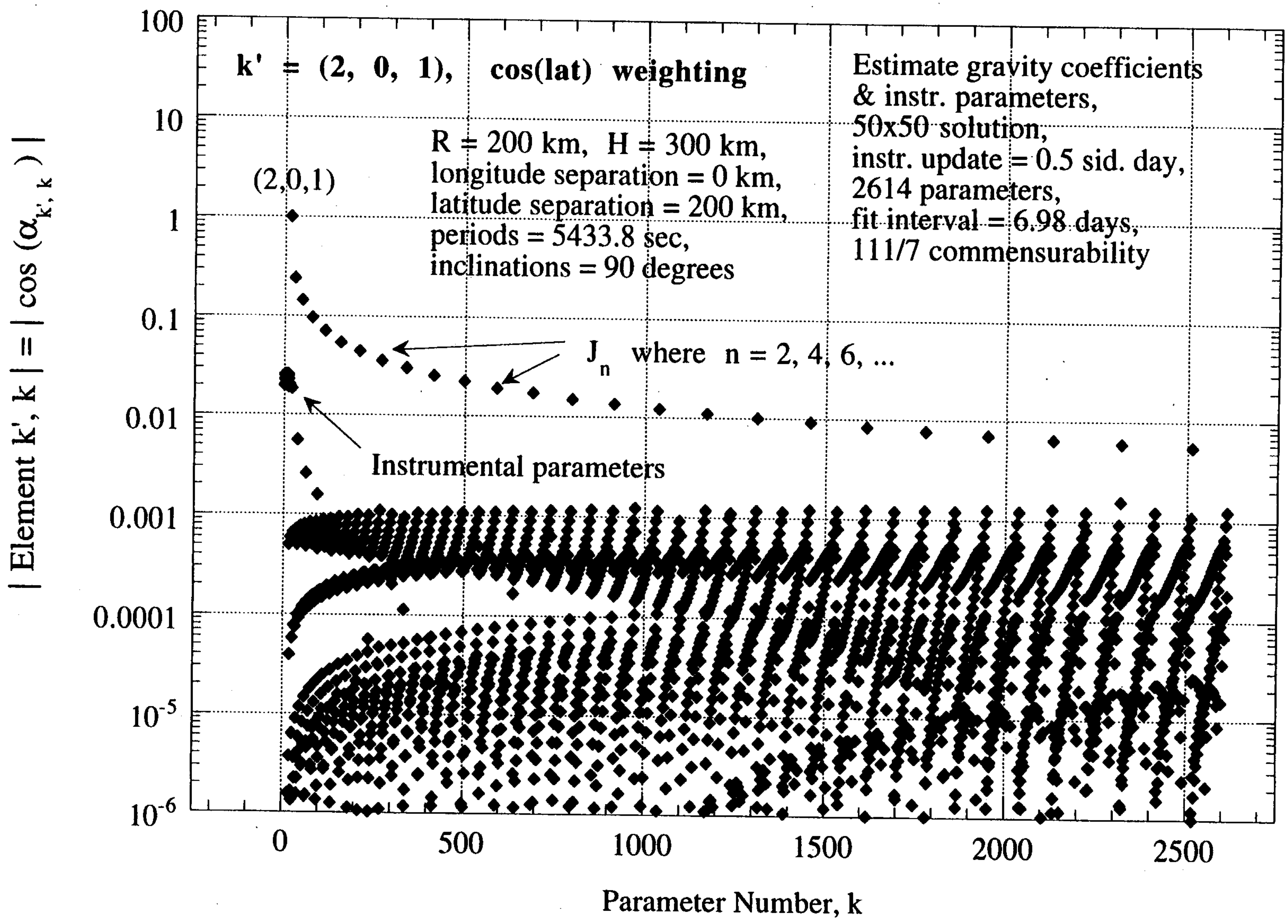

Figure D-9 Elements in Normalized A^TA, Row with $k' = (2, 0, 1) = J_2$, for the Corrected Acceleration Observable, with cos(lat) Weighting

178

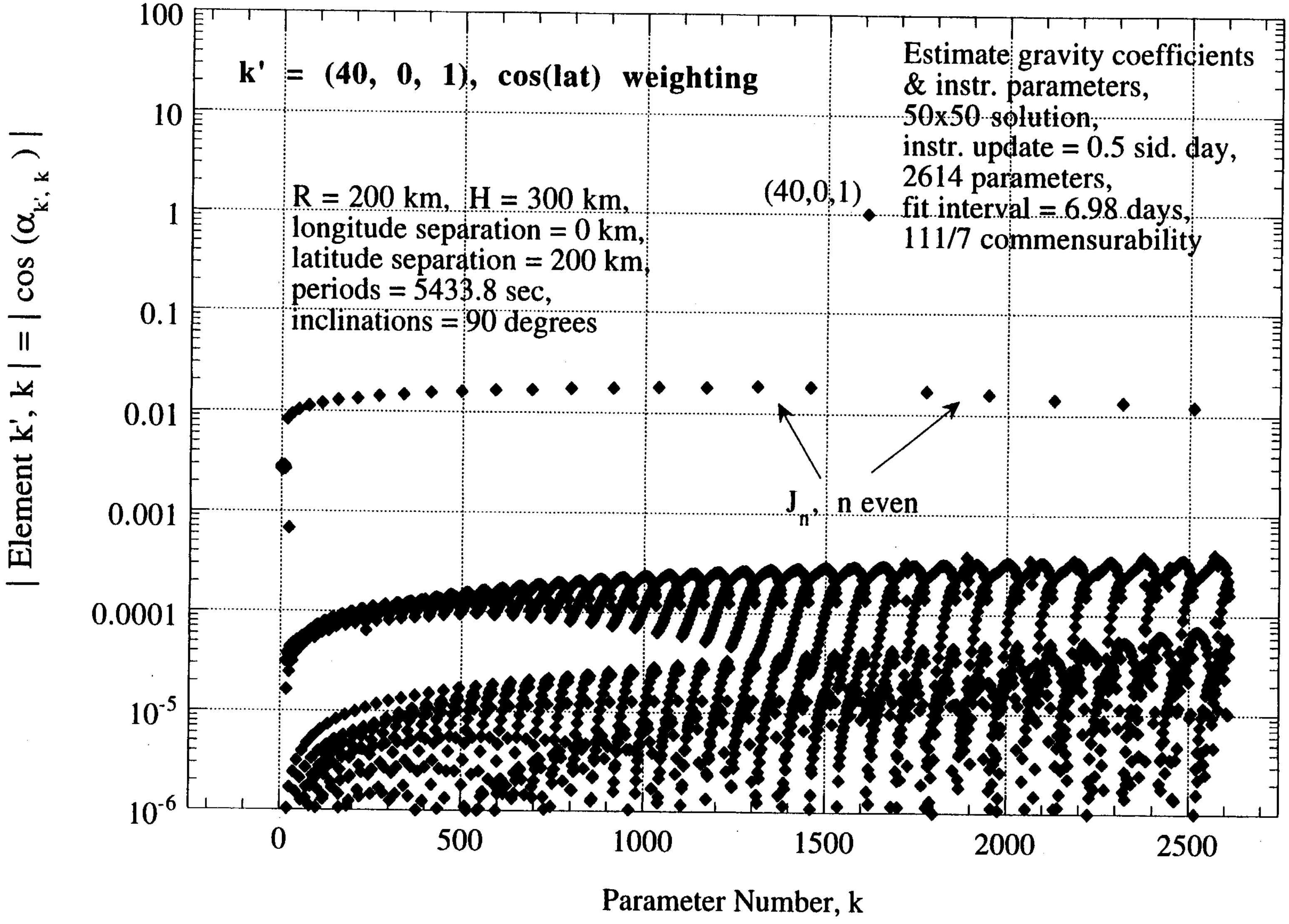

Figure D-10 Elements in Normalized A^TA, Row with k' = (40, 0, 1) = J_{40}, for the Corrected Acceleration Observable, with cos(lat) Weighting

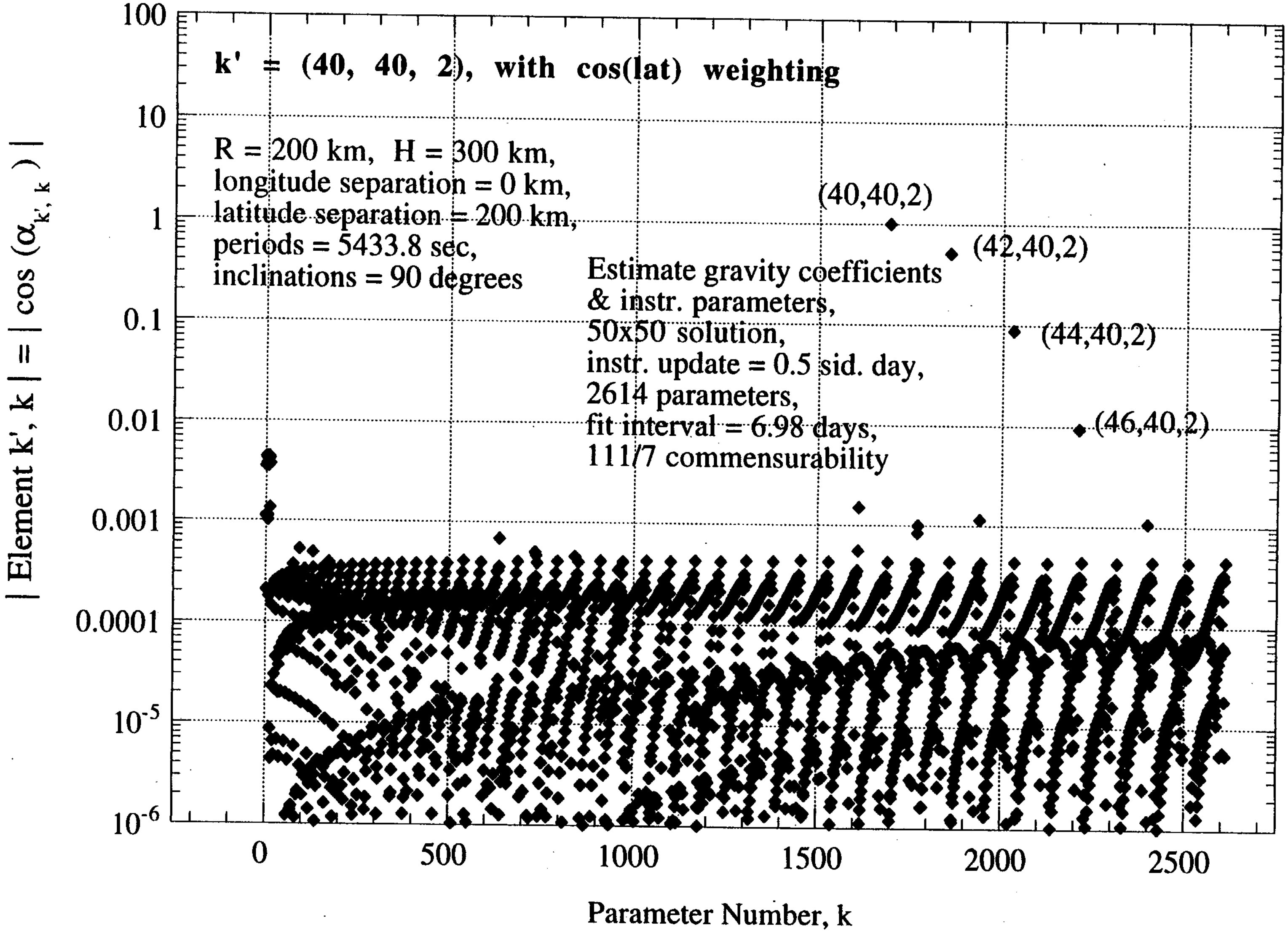

Figure D-11 Elements in Normalized A^TA, Row with k' = (40, 40, 2), for the Corrected Acceleration Observable, with cos(lat) Weighting

APPENDIX E

EXAMPLES OF LEGENDRE POLYNOMIALS

In order to provide some insight into their nature, this appendix presents examples of Legendre polynomials, and their first and second derivatives, for a number of order and degree values. The first derivative shows up in the gradient of the gravity field (i.e., acceleration) in the latitude direction. To first approximation (at least when the range between satellites is sufficiently small), the second derivative can show up in the difference in accelerations between satellites.

Figs. E-1, E-2 and E-3 present the normalized Legendre polynomials of (Legendre) order 2, 20, and 50, which correspond to zonal spherical harmonics of (degree, order) = (2, 0), (20, 0), and (50, 0), respectively. Multipliers are applied as needed to make the Legendre polynomial and its derivatives all plot well on the same figure. As degree increases, the number of zero crossings increases, with n zero crossings per π. All latitudes exhibit substantial amplitude. However, note the relatively large amplitudes that would be reached near the poles for the higher-degree zonals.

To demonstrate dependence on order, Figs. E-4 through E-9 present the normalized associated Legendre polynomials for (n, m) = (50, 1), (50, 10), (50, 25), (50, 40), and (50, 50). As order (m) increases, amplitude strength shifts more toward the equator and the number of zero crossings decreases. For (50, 50) in Fig. E-9, there are no zero crossings and strength is concentrated in a relatively narrow band of latitude. Based on half maxima, the band spans $\pm 9.5^{\circ}$ for this example. (This shape can be easily computed in (n, n) cases for which the polynomials are proportional to the simple form $[\cos(\mathrm{lat})]^{n}$.) For comparison with (50, 50), Figs. E-10 through E-12 present the associated Legendre polynomials for other selected "sectorials," namely (3, 3), (20, 20), and (100, 100). Note the equatorial band of strength becomes tighter as degree increases, reaching $\pm 6.7^{\circ}$ between half maxima for (100, 100).

In the corresponding spherical harmonics, these shapes are modulated in the longitude direction by $\cos(m\lambda)$ or $\sin(m\lambda)$ when $m \neq 0$, as indicated in Eq. (4.4). Thus, a sectorial, particularly for higher degree values, can be visualized as a modulated band circling the unit sphere at the equator, becoming narrower in latitude spread as degree increases (approximately in proportion to $1/\sqrt{n}$).

Fig. E-13 illustrates the effect of weighting the second derivative of the degree-50 Legendre polynomial with $\sqrt{\cos(\mathrm{lat})}$. The large excursions found at the poles in the unweighted case are flattened out by the weighting and the resulting function is nearly sinusoidal with 50 zero crossings per π, if one excludes the zeroes at $\theta = \pm\pi/2$ induced by $\sqrt{\cos(\mathrm{lat})}$. As indicated in the figure, the RMS is reduced in magnitude by approximately a factor of 1.7 as a result of $\sqrt{\cos(\mathrm{lat})}$ weighting. This plot indicates that the second derivative of a Legendre polynomial of higher degree n, when weighted by $\sqrt{\cos(\mathrm{lat})}$, is nearly sinusoidal with n zero crossings per π (i.e., a frequency of n/rev for the corresponding gravity partial). Because each degree value leads to a nearly sinusoidal function with different integer cyc/rev, the weighted integral of the product of second derivatives of Legendre polynomials with different degree values, over all latitude values, is nearly zero:

$$\int \frac{d^2P_n}{d\theta^2}\,\frac{d^2P_{n'}}{d\theta^2}\cos(\theta)\,d\theta \approx 0, \quad n \neq n' \tag{E.1}$$

This result is used in Appendix D to assess correlations between errors on gravity coefficients estimated on the basis of the corrected acceleration observable.

Corresponding plots can be generated for P_n and $dP_n/d\theta$ and those plots lead to similar results. The weighted integral for $P_n P_{n'}$ when $n \neq n'$ is expected to be exactly zero because of the orthogonality of Legendre polynomials. The cos(lat) weighting decreases the RMS values by slightly larger factors than for $d^2P_n/d\theta^2$ (for degree 50, by approximately a factor of 1.8 for $dP_n/d\theta$ and by approximately a factor of 2.2 for P_n,).

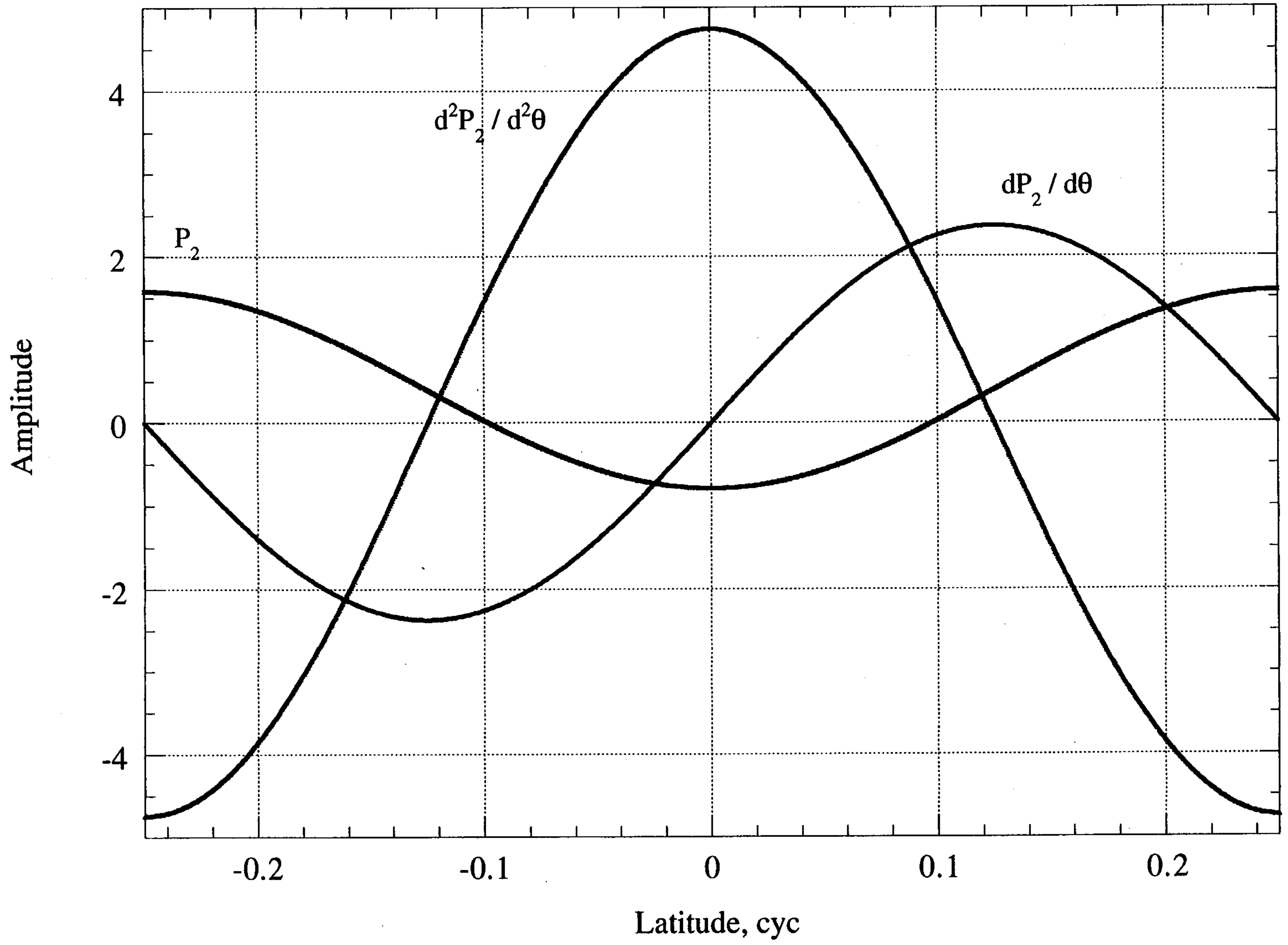

Figure E-1 Derivatives of the Normalized Legendre Polynomial of Order 2, with respect to Latitude

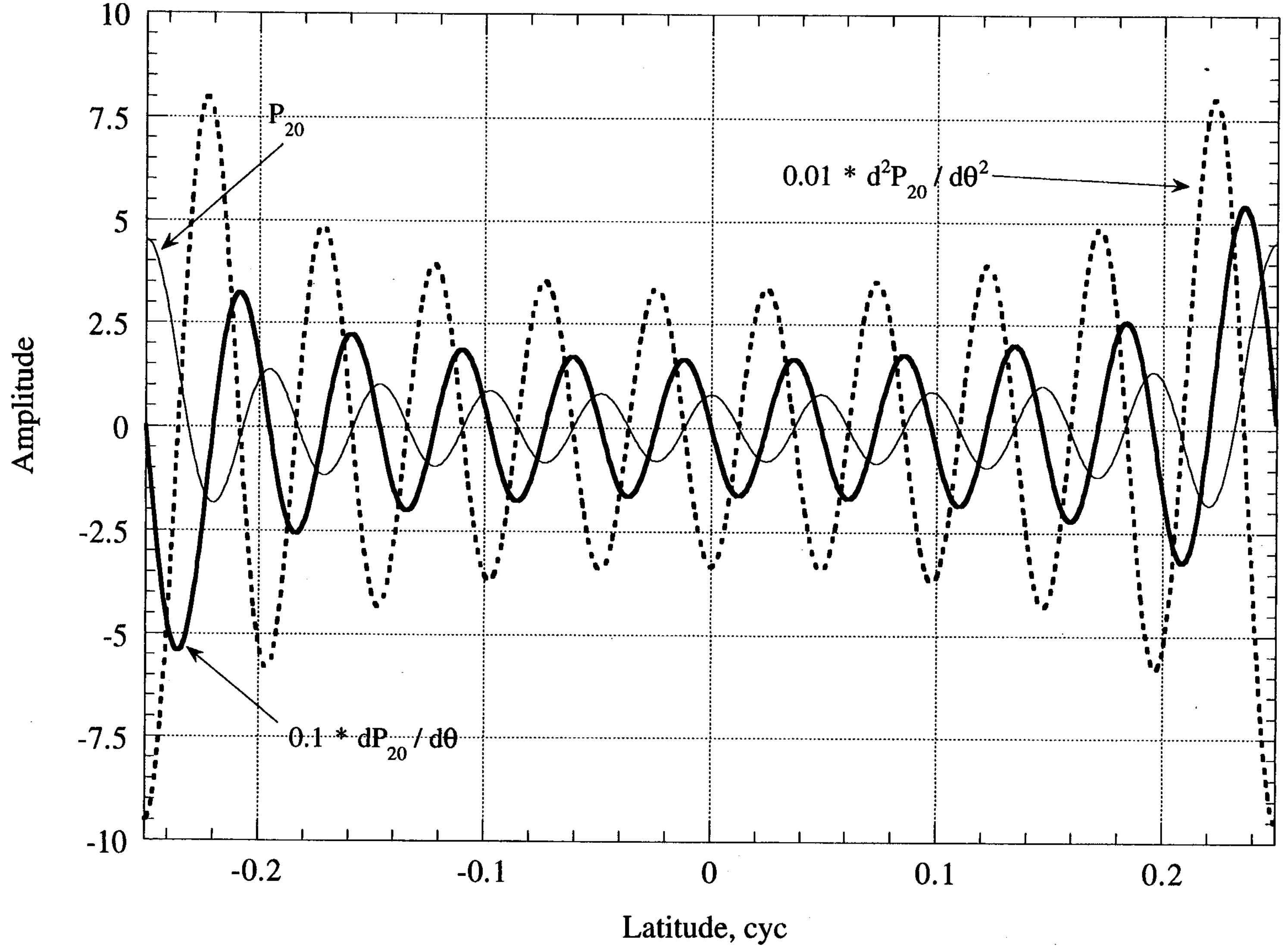

Figure E-2 Derivatives of the Normalized Legendre Polynomial of Order 20, with respect to Latitude

184

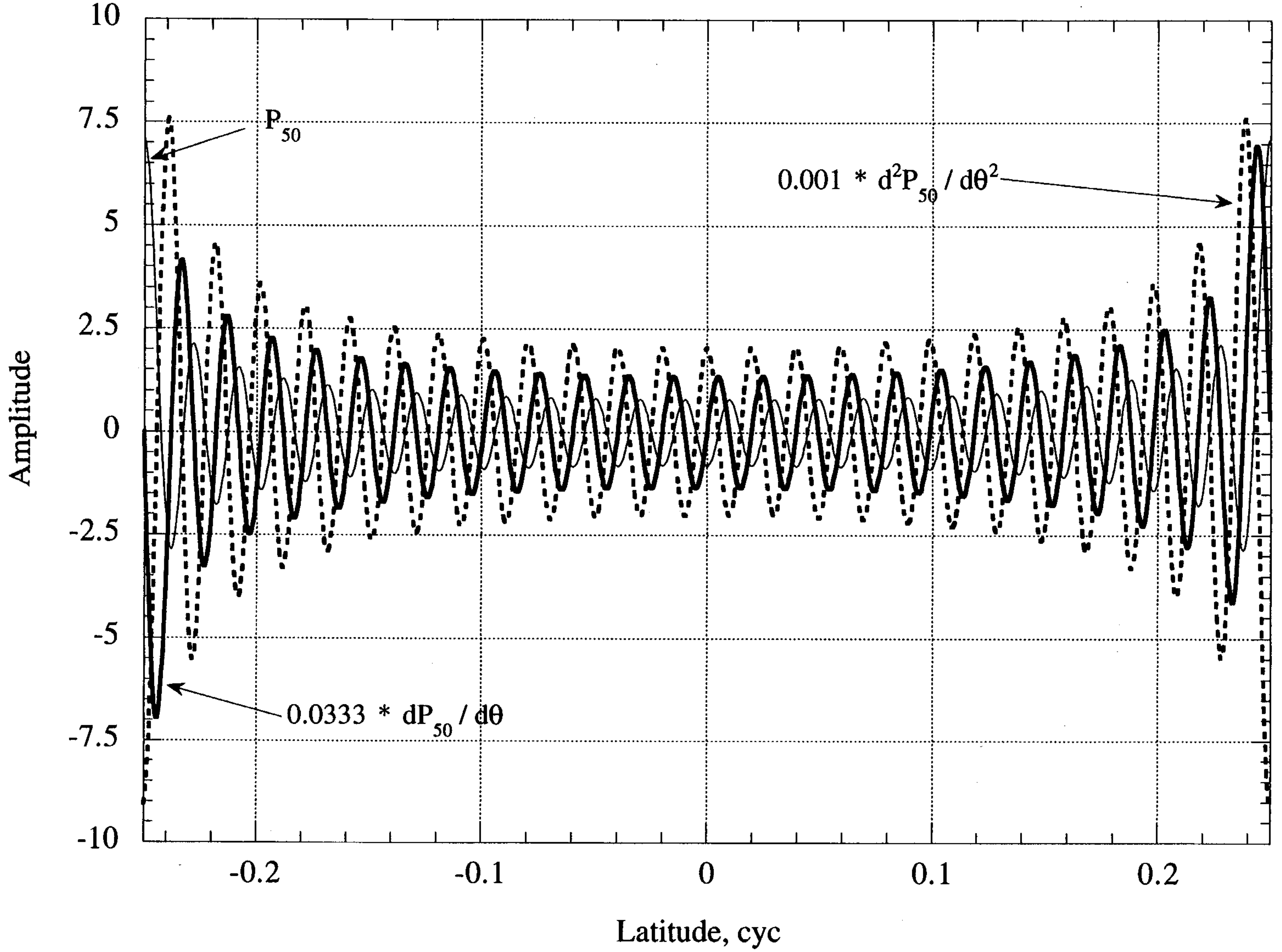

Figure E-3 Derivatives of the Normalized Legendre Polynomial of Order 50, with respect to Latitude

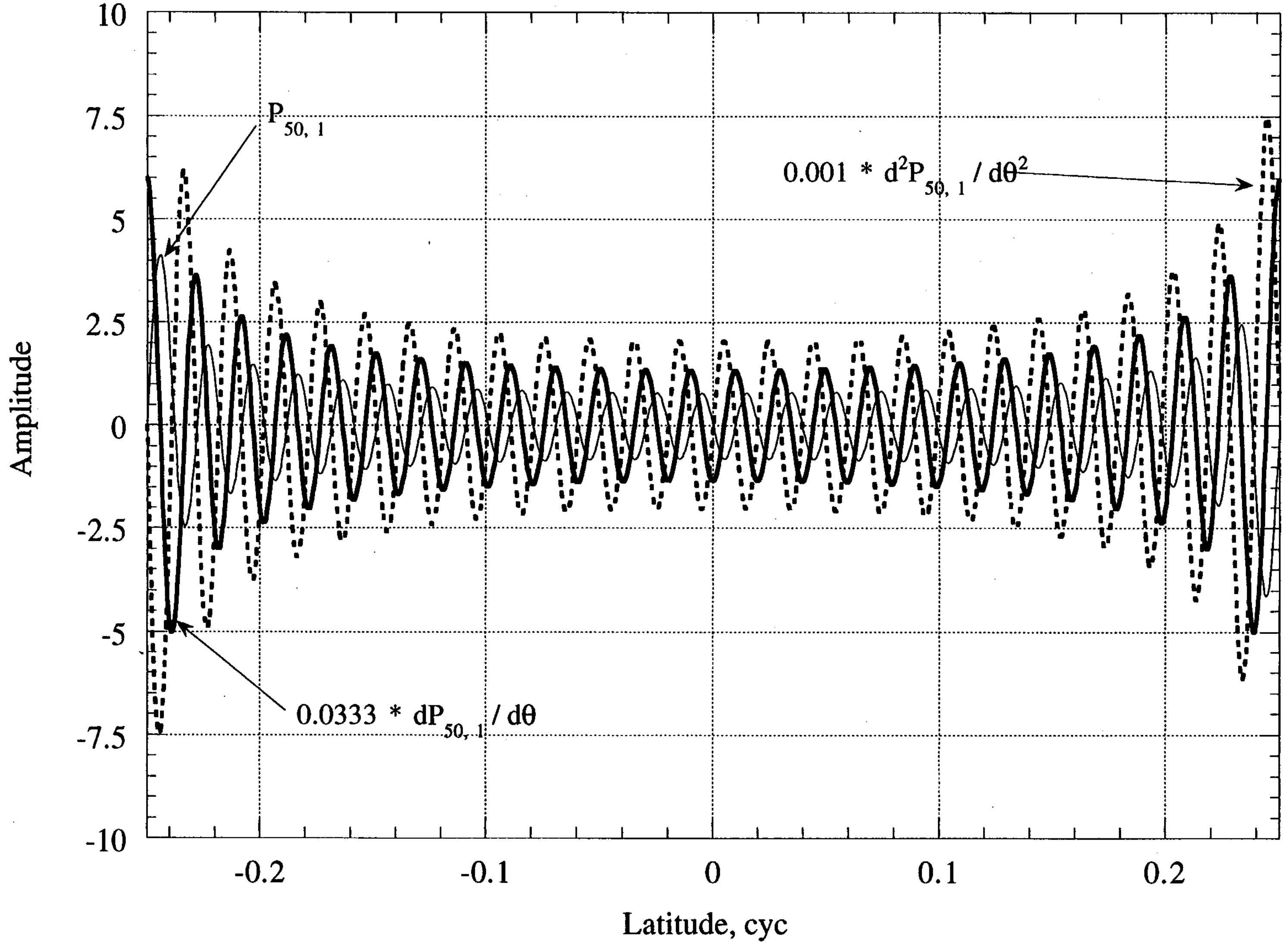

Figure E-4 Derivatives of the (50, 1) Normalized Associated Legendre Polynomial with respect to Latitude

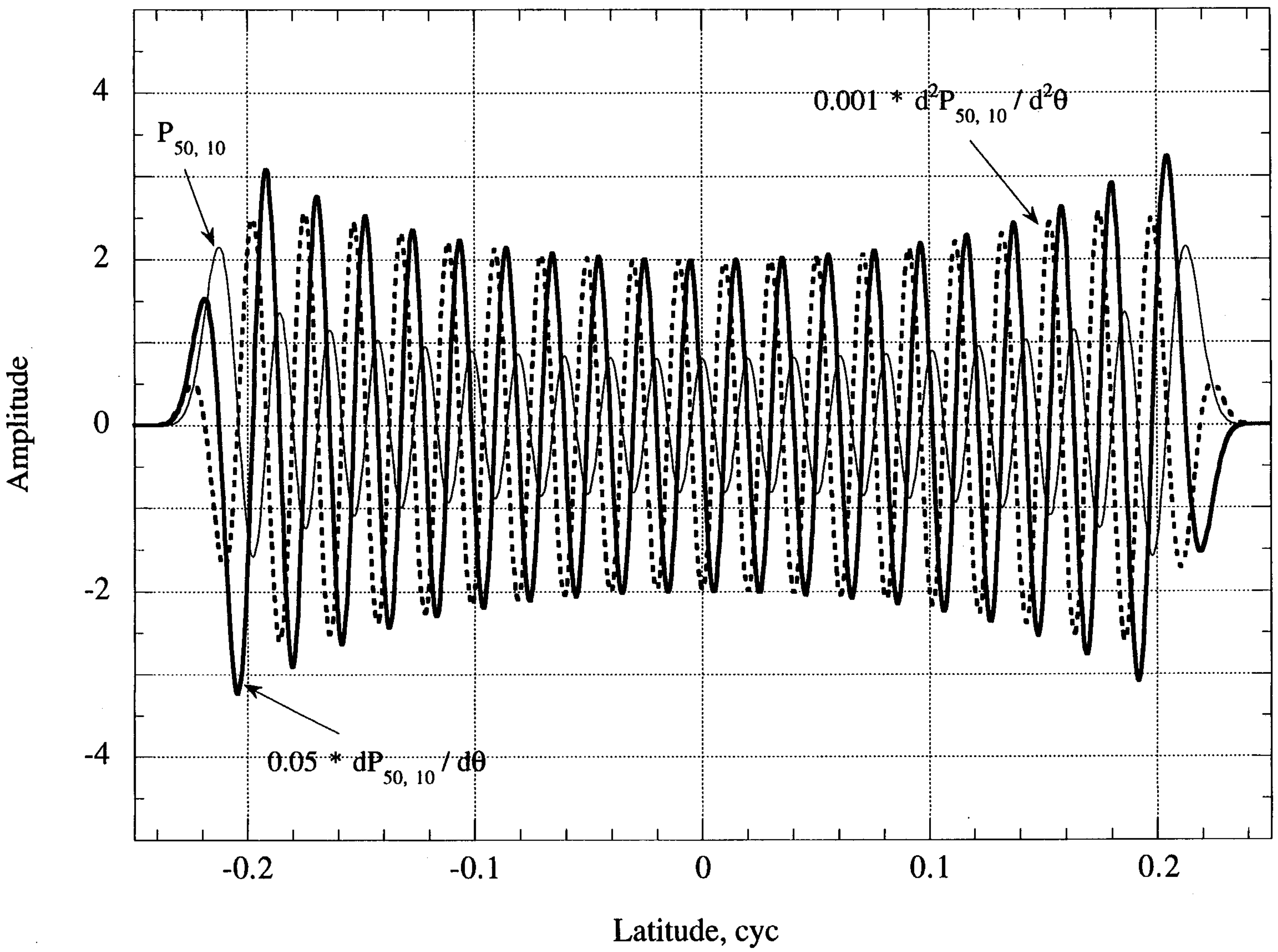

Figure E-5 Derivatives of the (50, 10) Normalized Associated Legendre Polynomial with respect to Latitude

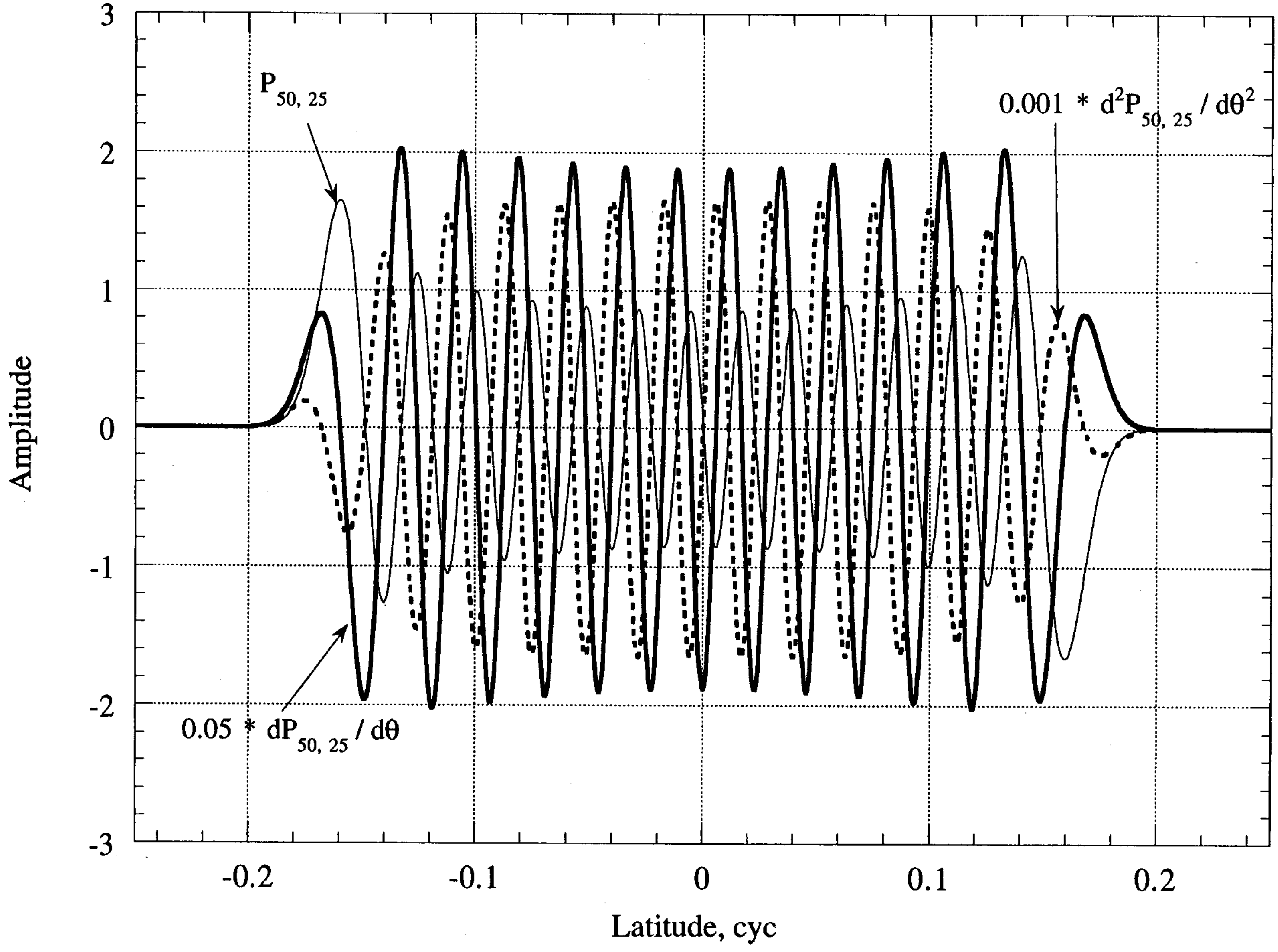

Figure E-6 Derivatives of the (50, 25) Normalized Associated Legendre Polynomial with respect to Latitude

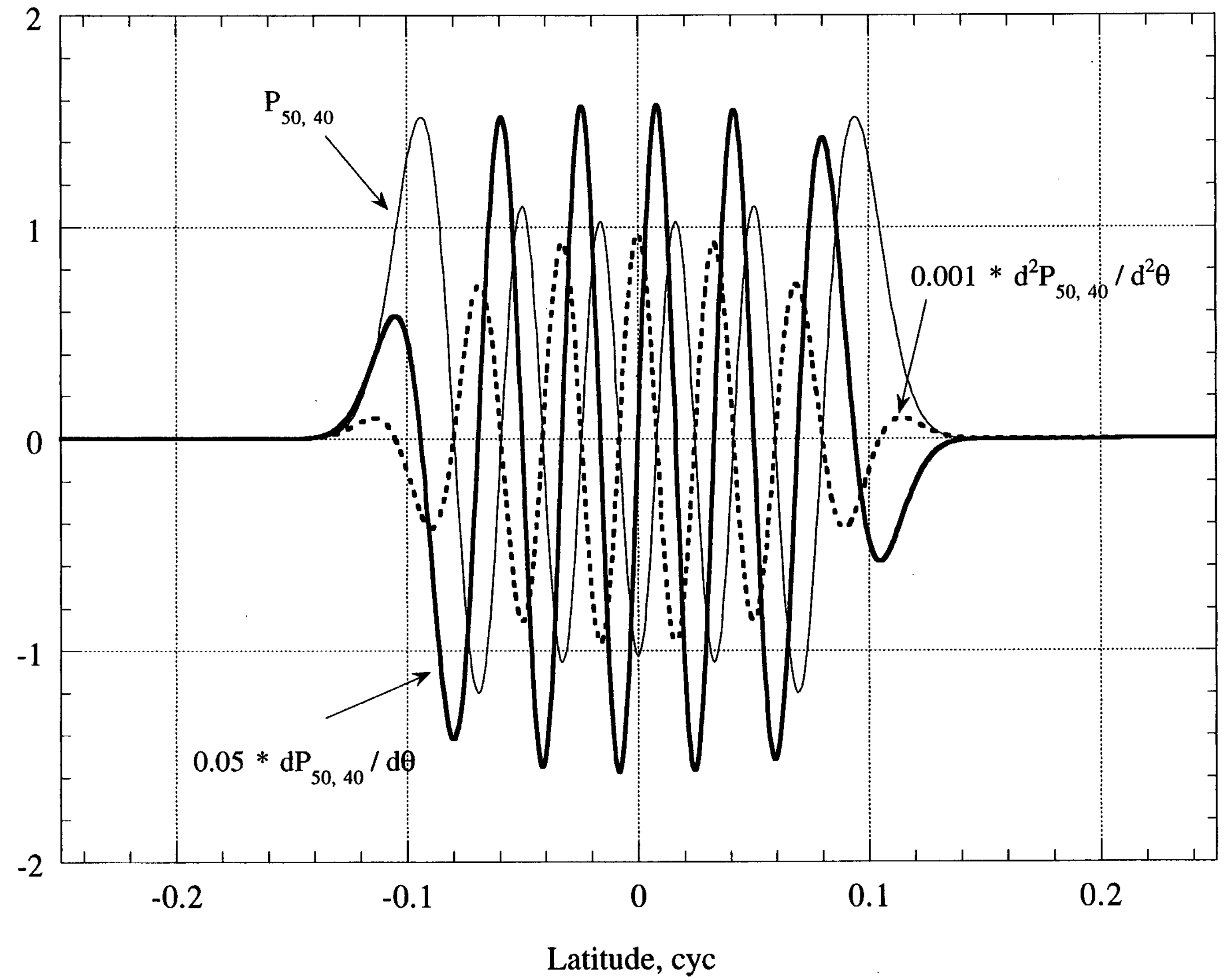

Figure E-7 Derivatives of the (50, 40) Normalized Associated Legendre Polynomial with respect to Latitude

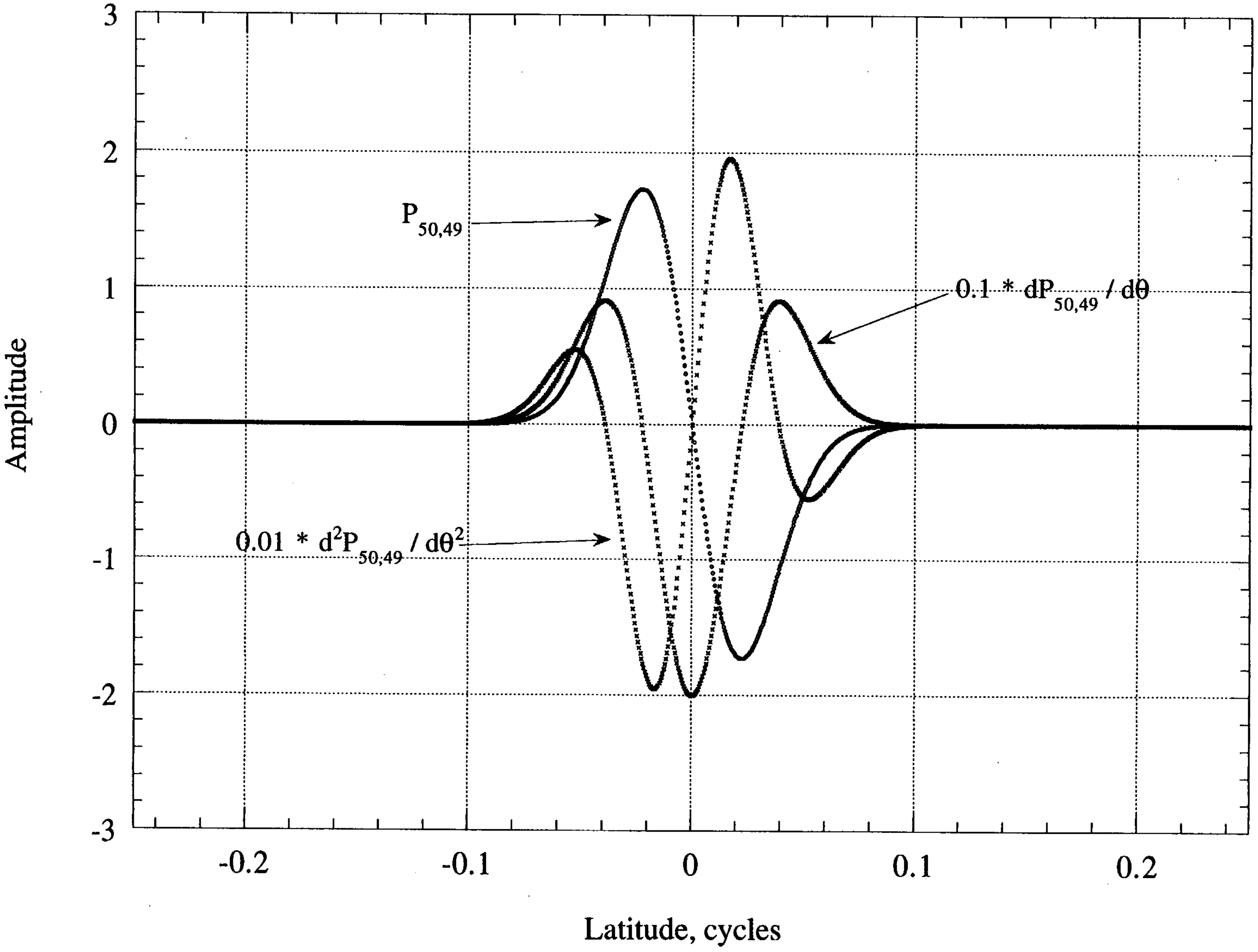

Figure E-8 Derivatives of the (50, 49) Normalized Associated Legendre Polynomial with respect to Latitude

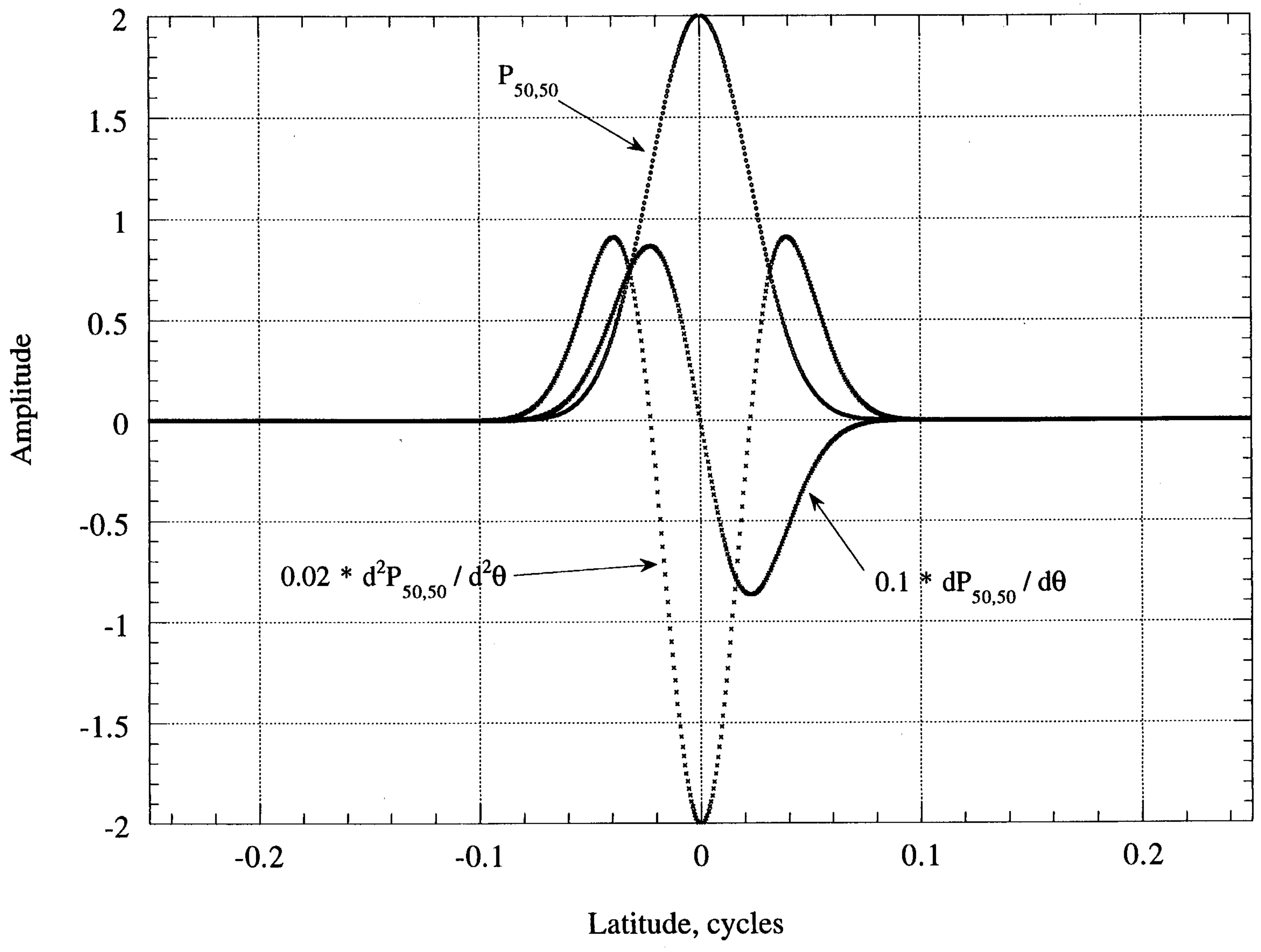

Figure E-9 Derivatives of the (50, 50) Normalized Associated Legendre Polynomial with respect to Latitude

191

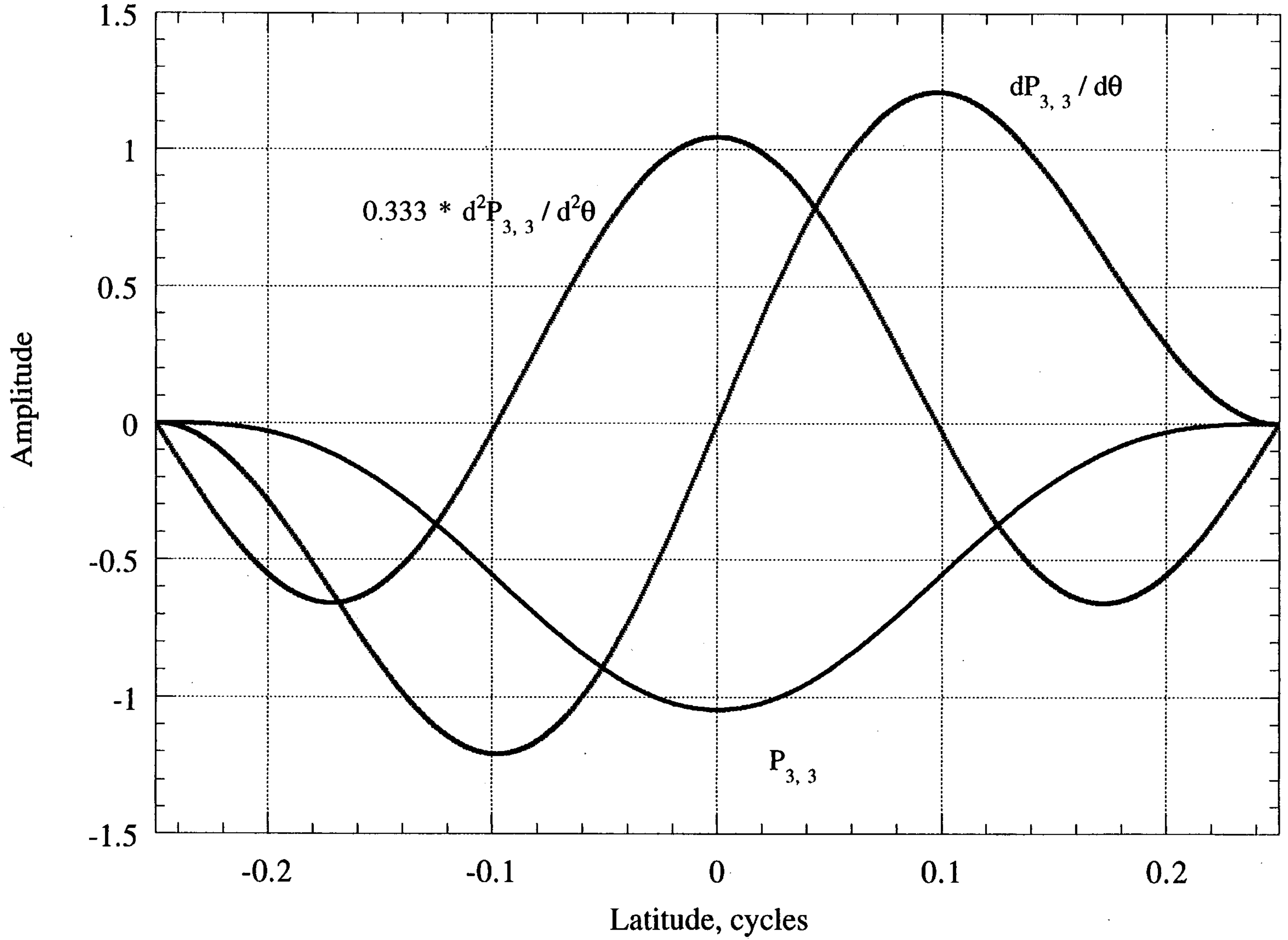

Figure E-10 Derivatives of the (3, 3) Normalized Associated Legendre Polynomial with respect to Latitude

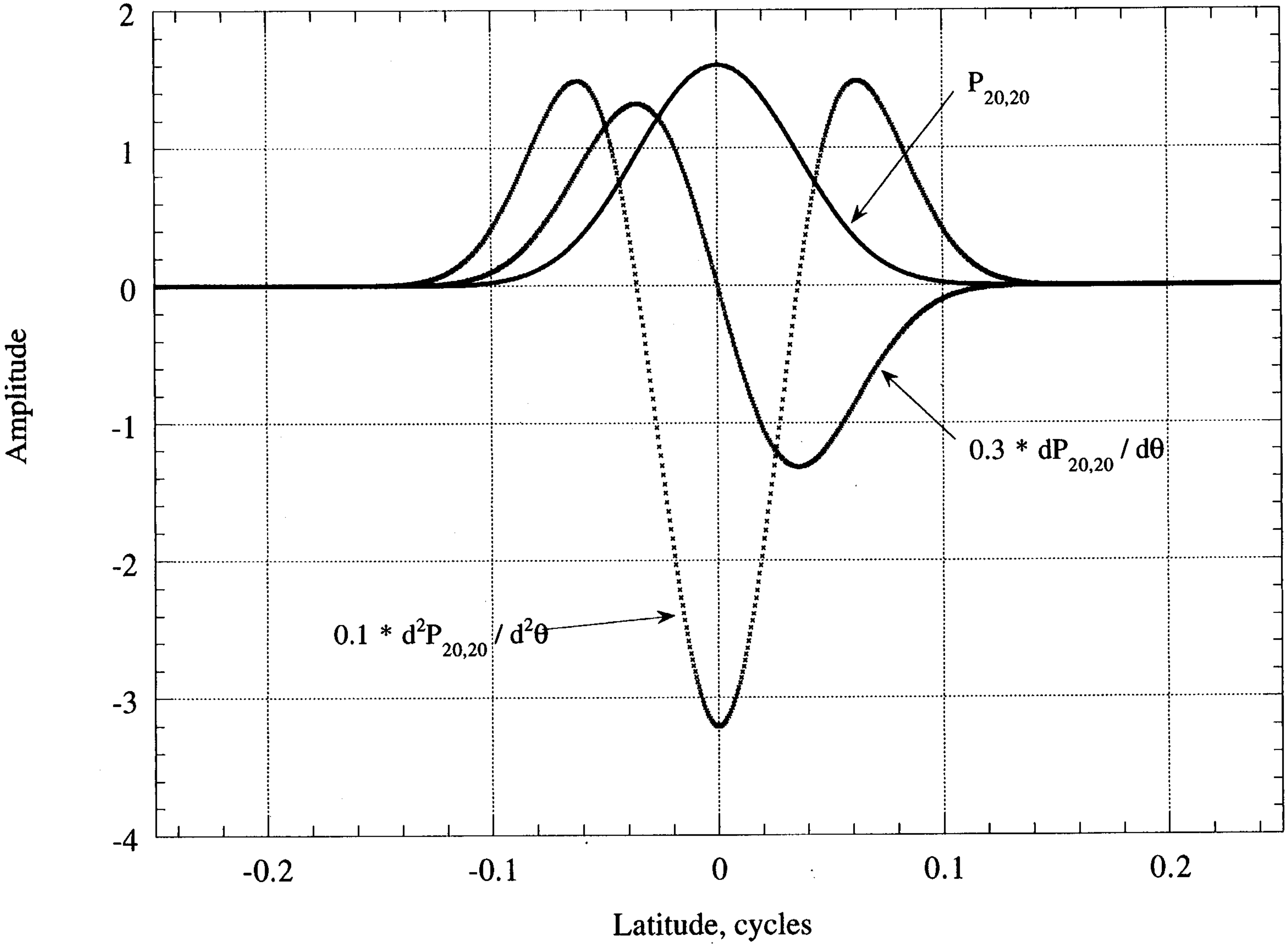

Figure E-11 Derivatives of the (20, 20) Normalized Associated Legendre Polynomial with respect to Latitude

193

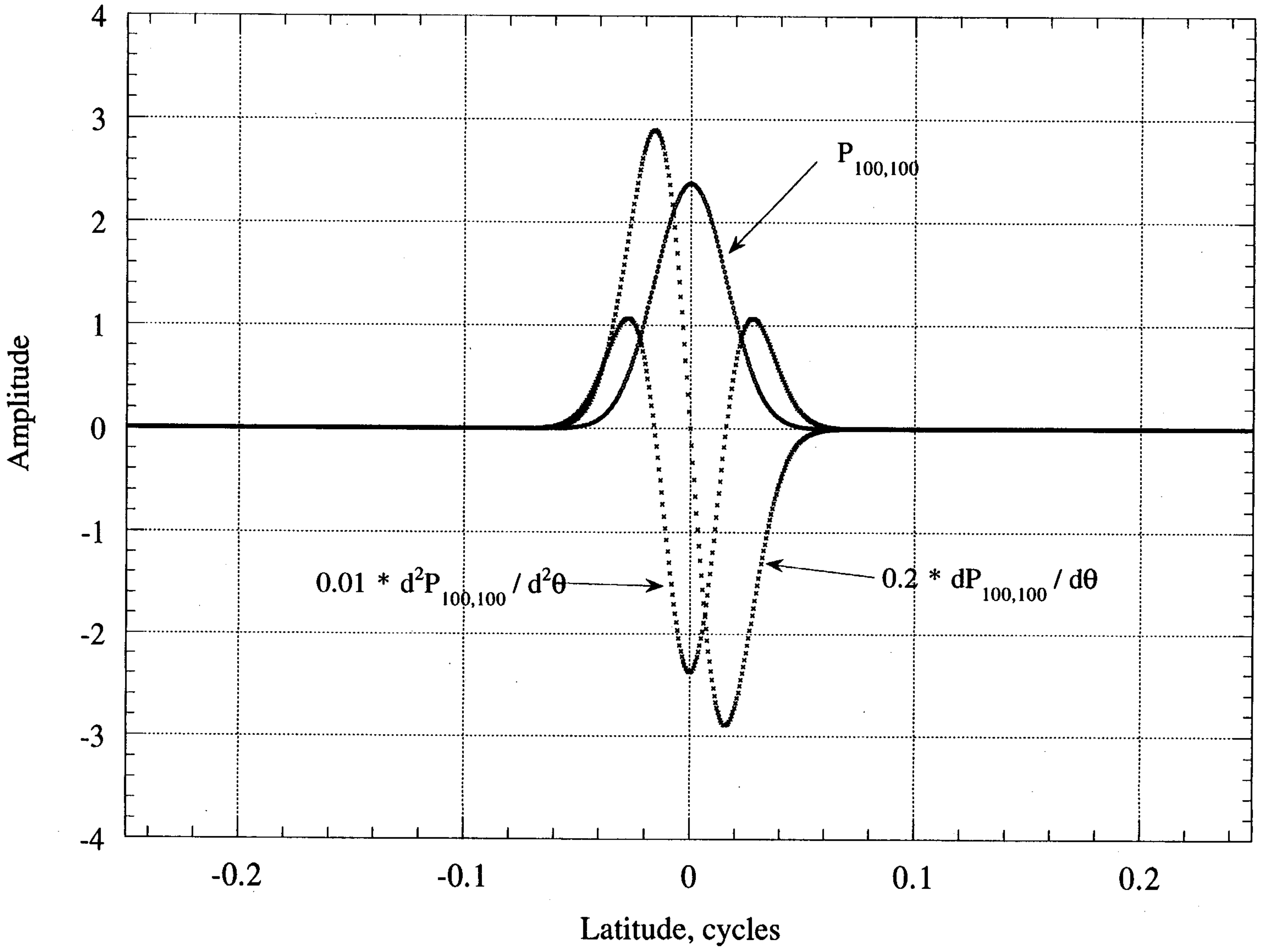

Figure E-12 Derivatives of the (100, 100) Normalized Associated Legendre Polynomial with respect to Latitude

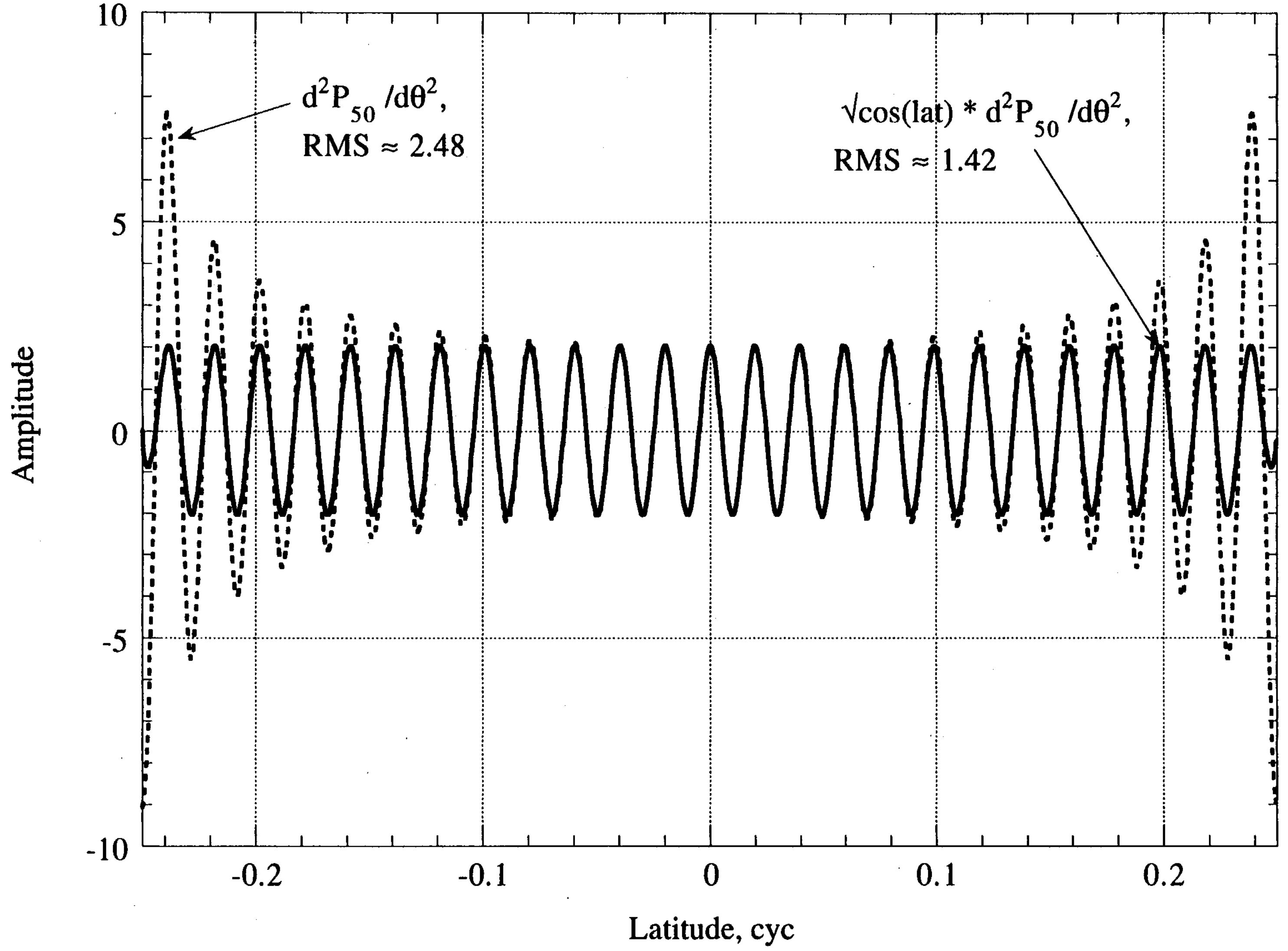

Figure E-13 Second Derivative of P_{50} with and without $\sqrt{\cos(\text{lat})}$ Weighting

REFERENCES

[1] MacArthur, J. L. and Posner, A. S., "Satellite-to-Satellite Range-Rate Measurement," *IEEE Trans. on Geoscience and Remote Sensing*, vol. GE-23, No. 4, pp 517-523, July, 1985.

[2] Sharma, J., "Precise Determination of the Geopotential with a Low-Low Satellite-to-Satellite Tracking Mission," Ph.D. Thesis, Univ. of Texas at Austin, 1995.

[3] Thomas, J. B., "An Analysis of Digital Phase-Locked Loops," JPL Publication 89-2, February, 1989.

[4] Stephens, S. A. and Thomas, J. B., "Controlled-Root Formulation for Digital Phase-locked Loops," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 31, No. 1, January, 1995, pp. 78-95.

[5] Kaula, W. M., *Theory of Satellite Geodesy*, Blaisdell Publishing: Waltham, MA, 1966

[6] Oppenheim, A. V. and Schafer, R. W., *Discrete-Time Signal Processing*, Prentice-Hall: Englewood Cliffs, N. J., 1989.

[7] Rohde, U. L., *Digital PLL Frequency Synthesizers*, Prentice-Hall: Englewood, NJ, 1983