

5-1993

Bayesian Estimation of Surface Information from Radar Images

Keith David Hartt

University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/open_access_dissertations



Part of the [Mathematics Commons](#)

Recommended Citation

Hartt, Keith David, "Bayesian Estimation of Surface Information from Radar Images" (1993). *Open Access Dissertations*. 363.
https://scholarworks.umass.edu/open_access_dissertations/363

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

**BAYESIAN ESTIMATION OF SURFACE INFORMATION
FROM RADAR IMAGES**

A Dissertation Presented

by

KEITH DAVID HARTT

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 1993

Department of Mathematics and Statistics

©Copyright by Keith David Hartt 1993

All Rights Reserved

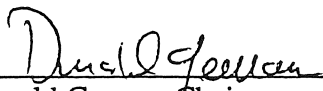
BAYESIAN ESTIMATION OF SURFACE INFORMATION
FROM RADAR IMAGES

A Dissertation Presented

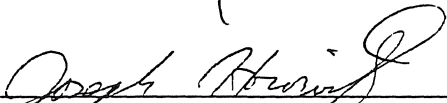
by

KEITH DAVID HARTT


Approved as to style and content by:



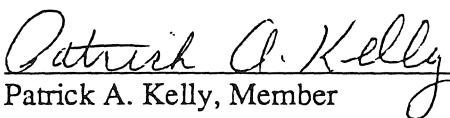
Donald Geman, Chair



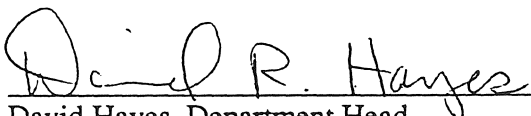
Joseph Horowitz, Member



Ramesh Korwar, Member



Patrick A. Kelly, Member



David Hayes, Department Head
Department of Mathematics and Statistics

ACKNOWLEDGEMENTS

The author gratefully acknowledges his advisor, Donald Geman, and his committee. He wishes to thank Ulf Grenander and Hans Künsch for helpful comments on the material in Chapter 4. He expresses his appreciation to James Head and Randolph Kirk for insights and help with the Magellan data. Thanks are due Basilas Gidas and George Reynolds for involvement at the outset of the work.

ABSTRACT

**BAYESIAN ESTIMATION OF SURFACE INFORMATION
FROM RADAR IMAGES**

MAY 1993

KEITH DAVID HARTT, B.S.E.E., UNIVERSITY OF MASSACHUSETTS

M.S.E.E., UNIVERSITY OF MASSACHUSETTS

Ph.D., UNIVERSITY OF MASSACHUSETTS

Directed by: Professor Donald Geman

The dissertation presents a method for deriving the shape of a surface from a radar image of the surface. An appropriate model of radar image formation is derived from physical principles. A Bayesian formulation of the inversion problem is developed upon which a computational strategy is based. Theoretical results on random surfaces relevant to the prior distribution are presented, and convergence and optimality properties of a new sampling algorithm are described. The technique is applied to Magellan data of Venus.

TABLE OF CONTENTS

	<u>Page</u>
ACKNOWLEDGMENTS.....	iv
ABSTRACT.....	v
LIST OF TABLES	viii
LIST OF FIGURES.....	ix
Chapter	
1. INTRODUCTION.....	1
Previous Work.....	2
Overview.....	2
Organization of Presentation.....	4
Contributions of the Thesis.....	4
References.....	5
2. SAR IMAGE FORMATION FOR RADARCLINOMETRY	6
Imaging Geometry.....	7
SAR Signal Processing.....	9
Surface Model.....	19
Speckle Statistics.....	21
Reflectance Function.....	27
Comparison to Standard Shape-from-shading	32
Examples	34
General Radiometric Model.....	35
Image Formation.....	36
References.....	37
3. RADARCLINOMETRY AS A STOCHASTIC INVERSE PROBLEM.....	38
Bayesian Formulation	38
Discrete Image Formation Model	42
References.....	44
4. REPRESENTATION OF STABILIZING PRIORS.....	45
Stabilizing Gibbs Distribution.....	45
Preliminaries.....	48
Representation.....	48
Properties.....	52
Standard Regularization Model.....	54

Self-similarity Property.....	63
References.....	71
5. EXTENDED GIBBS SAMPLER.....	72
Gibbs Sampler.....	73
Extended Gibbs Sampler.....	75
Applications.....	76
Back to Theorem 5.1.....	78
Extension.....	81
Optimality.....	82
Construction.....	85
References.....	86
6. RADARCLINOMETRY ALGORITHM.....	88
Posterior Distribution.....	88
Sampling Algorithm.....	89
Multi-resolution Scheme.....	91
Experiments.....	96
Discussion.....	115
References.....	115
BIBLIOGRAPHY.....	116

LIST OF TABLES

Table	Page
6.1. Flight parameters.....	97
6.2. Reconstruction parameters.....	98
6.3. Flight parameters.....	113

LIST OF FIGURES

Figure	Page
2.1. Imaging geometry	7
2.2. Top view of imaging geometry	8
2.3. Distance error functions	14
2.4. g_b functions.....	19
2.5. Surface model.....	20
2.6. Derivative of the sinc function.....	24
2.7. Sensing of a surface patch.....	29
2.8. Mesh plot of κ_r	32
2.9. Parallel projection	34
2.10. Flat plate.....	35
2.11. Cylindrical section.....	35
3.1. Penalty function.....	40
3.2. Reference surface for flat earth model.....	43
4.1. Sample from first order distribution.....	55
4.2. Sample from second order distribution.....	55
4.3. Sample from third order distribution.....	56
4.4. Sample from fourth order distribution	56
4.5. Mesh plot of first order sample	57
4.6. Mesh plot of second order sample.....	57
4.7. Mesh plot of third order sample.....	58
4.8. Samples from second order distribution.....	61
4.9. Central portions of Fig. 4.5	61
4.10. Mesh plot, periodic boundary.....	62
4.11. Mesh plot, free boundary	62
4.12. Difference variances.....	66

6.1. Sample evolution.....	94
6.2. Result of direct application of Gibbs sampler.....	95
6.3. Test surface	96
6.4. Noise-free radar image.....	97
6.5. 16-look radar image	97
6.6. Reconstructions.....	99
6.7. Noise-free image of reconstruction.....	99
6.8. Evolution of scale parameter estimate	100
6.9. Evolution of update ratio.....	100
6.10. Fixed-resolution reconstruction.....	101
6.11. Evolution of data energy	101
6.12. Sample from data distribution.....	102
6.13. Magellan radar image.....	103
6.14. Magellan elevation data	104
6.15. Radar image.....	105
6.16. Reconstructions.....	106
6.17. Noise-free image of reconstruction.....	106
6.18. Height image.....	107
6.19. Mesh plot of height image.....	107
6.20. Synthetic visual image	107
6.21. Evolution of scale parameter estimate	108
6.22. Evolution of update ratio.....	108
6.23. Radar image.....	109
6.24. Height image.....	109
6.25. Synthetic visual image	110
6.26. Magellan radar image.....	111
6.27. Synthetic visual image	112

6.28. Radar image.....	113
6.29. Reconstruction.....	113
6.30. Noise-free image of reconstruction.....	114
6.31. Height image.....	114
6.32. Synthetic visual image	114

CHAPTER 1

INTRODUCTION

This dissertation provides a technique for deriving surface properties from a radar image of the surface. The application area is remote sensing, as in the radar imaging of the surface of Venus by the Magellan probe. The goal is to derive from radar images important surface properties such as shape and dielectric attributes which are important to planetary geologists but not conveyed directly by the data. This is accomplished by posing the problem, known as *radarclinometry*, as a stochastic inverse problem.

Radar images acquired by synthetic aperture radar (SAR) represent different information than what ordinary visible-light images convey. For this reason radar images are not well suited for interpretation by the human visual system. The radar spectrum consists of lower frequencies relative to the visible spectrum; in addition, the imaging geometry is different. Radar images are acquired by air- or space-borne sensors which are “side looking”: a radar image represents energy returned as a function of flight-path position and range from the sensor, as opposed to light energy as a function of position on an image plane, which is the case for a standard image. An important benefit of radarclinometry is that a standard visible-light image, suitable for human interpretation, can be synthesized based on the surface shape the radarclinometry procedure generates. In the case of Magellan, the radar image data convey information unavailable from other sources, so radarclinometry is a valuable tool.

Previous Work

Techniques for radarclinometry, or radar shape-from-shading, are relatively rare in the literature. The first work in the area is due to Wildey [5], [6], [7] who poses the problem as the solution of a nonlinear first-order partial differential equation. Kirk [3]

uses a finite-element algorithm. Frankot and Chellappa [1] use shape-from-shading techniques. Guindon [2] and Thomas *et al.* [4] also present methods. As described later, there is a relationship between Bayesian methods and regularization problems. The techniques of Frankot and Chellappa and Thomas *et al.* pose radarclinometry as a regularization problem, but the resulting cost function is different from the one developed here.

A number of features distinguish the technique described here from previous work: 1) The formulation is Bayesian, i.e., the problem is posed as a stochastic inverse problem. 2) The height function is estimated directly, not through surface gradients. 3) Parallel projection is not assumed. 4) The data term is based on a speckle model. 5) The point spread function of the radar system is incorporated into the image formation process. 6) A realistic model is used for the surface, in that samples of the surface distribution are believable terrains.

Overview

The Bayesian model is derived from a physical model of the surface and a prototypical radar imaging system. The two components of the Bayesian approach are: 1) the image formation process, which provides the *data distribution*, the distribution of the radar data given the surface; 2) the *prior distribution*, a distribution placed on the surface shape which encodes prior knowledge about its statistical characteristics.

The SAR imaging system creates the radar image from the raw data: the radar pulses reflected back to the radar from the surface. Each value of the radar image represents energy returned from a subset of the surface, and the SAR signal processing localizes the energy from this ideally small subset. The image formation model, which gives the data distribution, has two components: 1) an image formation operator, which takes the surface attributes sought to a noise free version of the radar image; 2) the

distribution of the “speckle noise,” which is due to small-scale surface properties. Thus, the second component gives the distribution of the fluctuations about the expected energy specified by the first component.

The prior distribution is a Gibbs distribution with a second order Tikhonov stabilizer as the energy. In other words, the energy penalizes large second-order discrete partial derivatives. The choice of energy is motivated by regularization theory, but the distribution is attractive from the Bayesian perspective: samples from the distribution yield believable terrains.

Given the prior distribution and the data distribution, Bayes rule yields the *posterior distribution*, the distribution of the surface given the data. The computational method estimates the expected value of the surface given the radar image of the surface, the posterior mean.

As given here radarclinometry is similar to other stochastic inverse problems in image analysis. There is an image formation operator which takes attributes to image data. The attributes represent important information that is statistically related to the data but not directly conveyed by it.

The radarclinometry problem has large dimensionality, in that the height function, the surface attribute sought, is an image, a large data structure. This requires creativity in estimating the surface. The posterior mean cannot be computed directly as a weighted average, because the sum is too large to be computationally practical. Monte Carlo estimation is used instead, with a Markov chain (an extension of the Gibbs sampler) used to generate samples from the posterior distribution. Also, the estimation is imbedded in a multi-resolution hierarchy which gives rise to favorable convergence properties.

The computational procedure is most simply described as follows. Given a current surface state, a minor change is proposed to it. The radar image formation operator is applied to the resulting surface to obtain an ideal radar image based on the surface. This image is compared to the *observed* radar image. The change to the current state tends to be accepted if the comparison is favorable. The process is iterated, and in this way the surface evolves to a state which is consistent with the observations.

Organization of Presentation

The organization of the presentation is as follows. Chapter 2 covers the image formation process. The formulation of radarclinometry as a stochastic inverse problem is given in Chapter 3. Chapter 4 presents a set of results on a family of random surfaces, a member of which is used as the prior distribution of surface shape. Chapter 5 presents the extended Gibbs sampler, which is the Markov chain algorithm used to generate samples from the posterior distribution. Finally, Chapter 6 covers some final details of the radarclinometry algorithm and experimental results of its application to synthetic and Magellan Venus data.

Contributions of the Thesis

The novel aspects of the presentation are as follows. Chapter 2 represents a careful, comprehensive, and mathematically correct explanation of the radar image formation process. The result is a useful model for many applications, based on a prototypical radar imaging system and an explicit surface model. The main result of Chapter 4 is a representation theorem for a family of random surfaces based on using a Tikhonov stabilizer as the energy in a Gibbs distribution. The representation theorem allows derivation of : 1) the covariance structure of the field; 2) the normalizing constant for the distribution, important for obtaining the maximum likelihood estimator of a scale parameter; 3) a self-similarity property, which has been claimed to be appropriate in

modeling natural phenomena, and which also allows consistency of the prior among levels of the multiresolution hierarchy. Chapter 5 presents an extended Gibbs sampler used in the surface reconstruction procedure. Theoretical results are presented on convergence and optimality of the Markov chain. Also, special cases of the extended Gibbs sampler which are attractive for image reconstruction are presented. Chapter 6 presents a computational technique which can be applied to data such as that provided by the Magellan probe of Venus in order to obtain information on surface structure not available through other sources.

References

1. R. T. Frankot and R. Chellappa, (1990). "Estimation of surface topography from SAR imagery using shape from shading techniques," *Artificial Intelligence*, 43, pp. 271-310.
2. B. Guindon, (1989). "Development of a shape from shading technique for the extraction of topographic models from individual spaceborne SAR images," *Proc. of IGARSS 1989*, 2, pp. 249-263.
3. R. L. Kirk, (1987). "A fast finite-element algorithm for two-dimensional photogrammetry," Ph. D. thesis, California Institute of Technology.
4. J. Thomas, W. Kober, and F. Leberl, (1991), "Multiple image SAR shape-from-shading," *Photogrammetric Eng. Rem. Sens.*, 57, pp. 51-59.
5. R. L. Wildey, (1984). "Topography from single radar images," *Science*, 224, pp. 153-156.
6. R. L. Wildey, (1986). "Radarclinometry for the Venus radar mapper," *Photogrammetric Eng. Rem. Sens.*, 52, pp.41-50.
7. R. L. Wildey, (1988). "The surface integral approach to radarclinometry," *Earth, Moon, and Planets*, 41, pp. 141-153.

CHAPTER 2

SAR IMAGE FORMATION FOR RADARCLINOMETRY

This chapter describes the image formation process for a synthetic aperture radar image as is relevant to radarclinometry or radar shape-from-shading, the problem of deriving surface information from a radar image of the surface. A surface model is assumed which is valid for representing a useful class of terrestrial and planetary scenes. An important special case, used in the Magellan experiments presented later, assumes that surface composition is homogeneous so that the average radar energy reflected by surface patches depends only on orientation of the patch relative to the sensor. From the surface model the distribution of the radar image is derived, based on a realistic prototype SAR system. This establishes the correspondence between surface shape and the distribution of the radar image necessary in the formulation of radarclinometry as a stochastic inverse problem.

An overview of the presentation is as follows. First the sensor geometry is described. Next, the SAR signal processing mechanism is covered by deriving a point spread function associated with the conversion of received radar pulses to a radar image. The point spread function completely characterizes the SAR system signal processing to the extent that is required for rendering of images from the surface, i.e., implementing the forward imaging process. The physical surface model is presented next, from which the distribution of speckle, a phenomenon due to fine surface structure, is derived. Finally, the radiometric principles which govern the expected returned energy from the surface are presented.

Imaging Geometry

The geometry of the imaging process is as follows. A cylindrical coordinate system relative to the sensor is convenient, as depicted in Fig. 2.1.

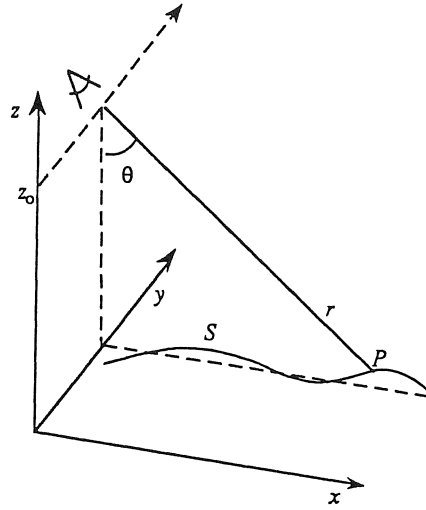


Fig. 2.1. Imaging geometry.

Surface location is given as a point set S . The sensor moves along a straight flight path (shown as a dotted line parallel to the y -axis) in the y -direction, known in radar literature as the *azimuth* direction, at constant velocity. Cylindrical coordinates are given by *range* $r = \sqrt{(z - z_0)^2 + x^2}$, and $\theta = \tan^{-1}(x/(z_0 - z))$, where z_0 is the height of the flight path, and x is the *ground range*. At regularly spaced positions $(r, y) = (0, n\Delta a)$, $n\Delta a \in I$, where the subset I of the real line is an interval, along the flight path a *pulse*, an electromagnetic signal of short duration, is transmitted by the sensor. Each pulse is assumed to illuminate the entire surface, and the echo of each pulse reflected by the surface, called a *return*, is collected and stored by the radar system. The pulses are spaced so that there is no overlap in the sense that each received pulse is due only to a single transmitted pulse. However, since each pulse illuminates the entire surface, each point on the surface contributes to all returns. A top view of the imaging geometry is given in Fig. 2.2.

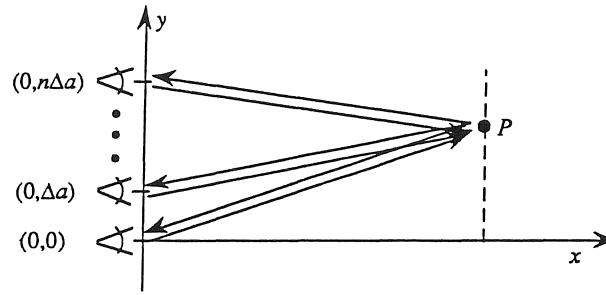


Fig. 2.2. Top view of imaging geometry.

The cylindrical coordinate system is convenient because a radar image is a function of range r and azimuth y , and for each (r, y) represents the energy returned by the set $S_{(r,y)} = \{(r', y', \theta): r' \approx r, y' \approx y, (r', y', \theta) \in S\}$. Thus the image at (r, y) is the energy returned from a subset of the surface. Referring to Fig. 2.1, for the purposes of data interpretation there is usually a correspondence made between the range coordinate r and estimated ground range x' given by $x' = \sqrt{r^2 - z_0^2}$. Clearly the true ground range is $x = \sqrt{r^2 - (z_0 - z)^2}$, so $x \neq x'$ unless $z = 0$ for all points in S . This is a distortion peculiar to radar images; the phenomenon of *layover* occurs when for two points $x'_1 < x'_2$ yet $x_1 \geq x_2$. The correct correspondence between ground range and range can be made if the surface shape is known, which is the purpose of this study. A *shadow* occurs in the radar image at (r, y) if $S_{(r,y)}$ is empty.

Surface location is reasonably described by means of a *surface location function* ρ , in that the surface point set is given as $S = \{(r, y, \theta): r = \rho(y, \theta), (y, \theta) \in G\}$, where $G \subset \{(y, \theta): y \geq 0, 0 \leq \theta \leq \pi/2\}$. For a fixed y and θ , the surface element first encountered along the corresponding ray, at a distance $\rho(y, \theta)$ from the flight path, is physically responsible for reflecting pulses. According to this principle, ρ is a well-defined function, i.e., not multi-valued. The radar image conveys no information about a surface facet that is hidden, so it is not represented by S . The definition of the surface location function is naturally suited to the imaging geometry. However, the relationship between the physical

properties of the surface and the surface location function is more complicated than the basics just portrayed; specifics are given later.

Summary. The radar image is a function of range and azimuth whose values represent the energy returned from $S_{(r,y)}$. The conversion of the returned pulses, one for each transmission location on the flight path, to the radar image is the province of the SAR signal processing, described in the next section. The changing position of the sensor relative to the surface allows what are essentially multiple views of the surface from different perspectives. The SAR signal processing combines the information from all returns (views) in the appropriate manner. As will be seen, the geometry of the situation gives rise to a relationship among the returns that can be exploited to localize the energy returned from the totality of S , represented in the raw returned pulses, to the energy returned from $S_{(r,y)}$, for each (r,y) . This localization is accomplished through a mechanism identical to the constructive and destructive interference of coherent light. Another analogy to optics is perpetuated by the terminology: the localization process afforded by the signal processing accomplishes what an aperture does, hence “synthetic aperture.” The details are as follows.

SAR Signal Processing

In this section the mechanism for conversion of received radar returns from the surface to a radar image is covered. A useful supplement to this development is Munson and Visentin [9], because in this presentation the same viewpoint is taken, in that the radar system is treated in a direct manner as a linear system, avoiding the confusing but prevalent “doppler” terminology. The aforementioned paper provides some system-specific detail not presented here, and contains references to the vast literature on SAR. The next paragraph is preliminary.

Signal Representation. A complex-valued signal (function of time) x has an associated signal

$$z(t) = x(t) \exp i 2 \pi f_0 t \quad (2.1)$$

where f_0 is a positive constant, known as the microwave *carrier frequency*, which is, up to an order of magnitude, 10^6 /sec. Under reasonable conditions (see Franks [4]) there is a direct correspondence between z and a physical waveform, from which the signal x , called the *complex envelope* of z , can be obtained. The quantity $|x(t)|^2$ is referred to as the *energy* of $x(t)$ (or of $z(t)$), and $\arg x(t)$ is sometimes called the *phase*.

System Model. Let the complex envelope of the transmitted radar pulse be denoted by $p(t)$, where $t = 0$ corresponds to the time at which each pulse is transmitted; i.e., the time reference is reset at each pulse transmission. Let the surface consist of a point scatterer at (r', y', θ) , ideal in the sense that it is a point reflector which returns the transmitted radar pulse without distorting the waveform, i.e., the received signal for the n^{th} sensing is simply a delayed version of the transmitted signal $p(t) \exp i 2 \pi f_0 t$, so the complex envelope $h_n(t, r', y')$ of the n^{th} received signal is given by

$$h_n(t, r', y') \exp i 2 \pi f_0 t = p(t - \tau_n(r', y')) \exp i 2 \pi f_0 (t - \tau_n(r', y')) \quad (2.2)$$

with

$$\tau_n(r', y') = \frac{2}{c} \sqrt{(y' - n \Delta a)^2 + r'^2} \quad (2.3)$$

where c is the speed of light and $\tau_n(r', y')$ is simply the two-way time delay between the points $(0, n \Delta a, \theta)$ and (r', y', θ) . The left hand side of (2.2) corresponds to a physical waveform. Note $h_n(t, r', y')$ does not depend on θ because the distance calculation does not involve it.

Note. As a practical matter, it has been tacitly assumed that the speed of the sensor platform is negligible relative to the speed of light, so that the sensor location does not change between the times of signal transmission and reception.

Now invoke the principle of superposition. Let the set of complex envelopes of the returns from an arbitrary surface be given by $\{u_n(t): n\Delta a \in I\}$, i.e., a set of complex-valued functions of time, where for each return $t = 0$ corresponds to the time at which the signal was transmitted. The returns are given by

$$u_n(t) = \iiint a(r', y', \theta) h_n(t, r', y') r' d\theta dr' dy' \quad (2.4)$$

where a , a characteristic of the surface, is a complex valued function which weights the ideal response assumed by h .

Note. The representation (2.4) is standard in the radar literature. However, as will be evident when the complex reflectance function a is defined later, the representation (2.4) is only symbolic, as in this presentation the integral reduces to a sum (see (2.23)).

The SAR processing is a linear operation, given by

$$Z(r, y) = \sum_{n \in A(y)} \int u_n(t) w_n(t, r, y) dt \quad (2.5)$$

where Z is the complex-valued radar image, and $w_n(t, r, y)$ for fixed (r, y) is a complex-valued weighting function. The radar image is random because the returns are random, as described later. The sum is over the set $A(y) = \{n: (y - n\Delta a) \in [-B/2, B/2]\}$, where B determines an interval in azimuth. Ideally, the weighting function is constructed so it selects the portion of each return $u_n(t)$ which corresponds to the energy reflected by $S_{(r, y)}$, in so doing *resolving* the surface, or scene.

Note. As a practical matter, B is chosen so that the antenna response is constant over the interval, and so that the straight-line flight path geometry is valid. The sum in (2.5) is what characterizes SAR relative to standard radar, giving improved resolution in azimuth.

Combining (2.4) and (2.5) and simplifying, notably changing orders of integration

$$Z(r, y) = \iiint a(r', y', \theta) k(r, y, r', y') r' d\theta dr' dy' \quad (2.6)$$

where

$$k(r, y, r', y') = \sum_{n \in A(y)} \int h_n(t, r', y') w_n(t, r, y) dt \quad (2.7)$$

is the *point spread function*, the output of the SAR processing at (r, y) due to an ideal point scatterer at (r', y', θ) .

The above (2.6) shows the sufficiency of k for characterizing the SAR system.

Recall that a is a function, characteristic of the surface, taking complex values which are reflection coefficients, in that these values modify the ideal reflectance assumed by h . The radar image should estimate $\int \int_{r' \approx r, y' \approx y} a(r', y', \theta) r' d\theta dr' dy'$ (related to $S_{(r, y)}$), so one seeks a

function k which is impulsive, i.e., takes its largest values when $y \approx y'$ and $r \approx r'$. Further, k is determined by p and w . The important issue is that the point spread function is essentially *designed* (to have desirable properties) through the choice of the functional form of the illuminating pulse and specification of the way in which returns are processed. This design is aided by the theory of *radar ambiguity functions*, which is described in Franks [4].

Point Spread Function. In the following the point spread function, the response of the radar system to an ideal scatterer, for a standard SAR system is derived (by specifying the standard choices for p and w and computing an approximation to (2.7)). Although only a prototypical situation is addressed (mainly for simplicity), the point spread

function is valid for a wide variety of systems such as Magellan. Further, the basic formulation is applicable in more complex situations: any system can be characterized by deriving the response to ideal elements.

The standard pulse is a “chirp.” Its complex envelope is

$$p(t) = \exp(i\pi\gamma t^2) \mathbb{1}_{[-T/2, T/2]}(t) \quad (2.8)$$

where γ is a real, positive constant, and T is the pulse width. Let

$E \subset (0, \infty) \times (-\infty, \infty) \times [0, \pi/2)$ be a compact set, known *a priori* to contain S , and let the domain of the radar image be a finite set F . Clearly for any sensible system $F \subset \{(r, y): (r, y, \theta) \in E\}$. In typical situations

$$\sup\{|(y' - n\Delta a)/r'|: (r', y', \theta) \in E, n\Delta a \in I\}$$

is small, so that the use of the Taylor expansion $(1 + x^2)^{1/2} = 1 + \frac{1}{2}x^2 + O(x^4)$ is justified on (2.3), and the time delay

$$\begin{aligned} \tau_n(r', y') &= \frac{2}{c} r' (1 + e_1((y' - n\Delta a)/r')) \\ \tau_n(r', y') &= \frac{2}{c} r' \left(1 + \frac{1}{2} ((y' - n\Delta a)/r')^2 + e_2((y' - n\Delta a)/r') \right) \end{aligned} \quad (2.9)$$

where errors $e_1(x) = (1 + x^2)^{1/2} - 1$, and $e_2(x) = (1 + x^2)^{1/2} - (1 + \frac{1}{2}x^2)$. It is easy to verify that $|e_1|, |e_2|$ increase with $|x|$. Plots are shown in Fig. 2.3, which shows the second approximation of (2.9) to be significantly better than the first. With regard to the specification of $w_n(t, r, y)$, the expansions (2.9) for τ_n are used to approximate the *matched filter* (which has some optimality properties, see Franks [4]), which requires $w_n(t, r, y) \propto \bar{h}_n(t, r, y)$ (the bar denotes complex conjugation), yielding (see (2.2))

$$w_n(t, r, y) = \frac{1}{T} \bar{p}(t - \tau_n(r, y)) \frac{1}{\# A(y)} \exp i 2\pi f_0 \tau_n(r, y) \quad (2.10)$$

The function is normalized so that for any (r, y) , $k(r, y, r, y) = 1$.¹ Instead of (2.10), the matched filter is approximated using the expansions (2.9) for τ_n ; the weighting function is defined by

$$w_n(t, r, y) = \frac{1}{T} \bar{p}\left(t - \frac{2r}{c}\right) \frac{1}{\#A(y)} \exp\left(i2\pi f_o \frac{2}{c} \left(r + \frac{1}{2}(y - n\Delta a)^2 / r\right)\right) \quad (2.11)$$

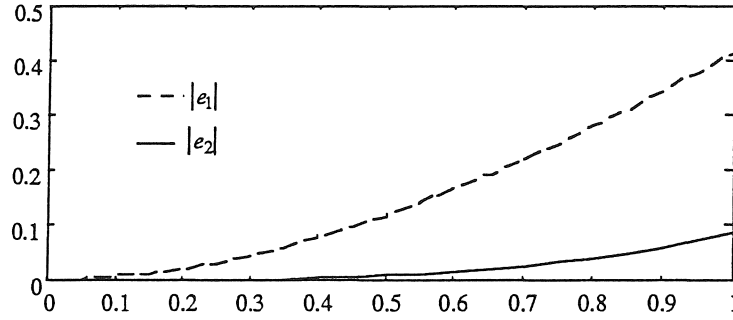


Fig. 2.3. Distance error functions.

Let $g_b(x) = (1 - |x|/b) \text{sinc}((1 - |x|/b)x) 1_{[-b, b]}(x)$, with $\text{sinc}(x) = \sin(\pi x)/\pi x$, and

$$\phi(r, y, r', y') = \frac{2\pi}{\lambda_o} (2(r - r') - (y - y')^2 / r) \quad (2.12)$$

with $\lambda_o = c/f_o$, the wavelength of the carrier.

Theorem 2.1. $\left| k(r, y, r', y') - g_{\gamma T^2} \left(\frac{r - r'}{c/2\gamma T} \right) g_{\infty} \left(\frac{y - y'}{r\lambda_o/2B} \right) \exp i\phi(r, y, r', y') \right| \rightarrow 0$ as $r \rightarrow \infty$,

$r' \rightarrow \infty$ and $\Delta a \rightarrow 0$.

Note. The limit conditions correspond to moving the set E farther and farther from the flight path and letting the separation between sensings get closer together.

¹Hence k is not a point spread function in the usual sense (without further normalization).

The proof is deferred until after some preliminaries are proved. Let $\delta_0 = \sup_n |(y' - n\Delta a)/r'|$. Establish bounds for the approximations (2.9) as follows.

$$\begin{aligned}\delta_1 &= \sup_n \left| \tau_n(r', y') - \frac{2r'}{c} \right| \leq \frac{2r'}{c} \sup \{ |e_1(u)| : |u| \leq \delta_0 \} = \frac{2r'}{c} \left((1 + \delta_0^2)^{1/2} - 1 \right) = r' O(\delta_0^2) \\ \delta_2 &= \sup_n \left| \tau_n(r', y') - \left(\frac{2}{c} \left(r + \frac{1}{2} \frac{(y' - n\Delta a)^2}{r'} \right) \right) \right| \\ &\leq \frac{2r'}{c} \sup \{ |e_2(u)| : |u| \leq \delta_0 \} = \frac{2r'}{c} \left((1 + \delta_0^2)^{1/2} - \left(1 + \frac{1}{2} \delta_0^2 \right) \right) = r' O(\delta_0^4)\end{aligned}\tag{2.13}$$

Lemma 2.1. $\left| \frac{1}{T} \int (p(t - \tau_n(r', y')) - p(t - 2r'/c)) \bar{p}(t - 2r'/c) dt \right| \rightarrow 0$ as $r' \rightarrow \infty$.

Proof. First establish a bound for $|p(t - \tau_n(r', y')) - p(t - 2r'/c)|$. Let $|t - \tau_n(r', y')| \leq T/2$, $|t - 2r'/c| \leq T/2$, and let $v = t - 2r'/c$. Then using the basic inequality $|\exp(ix) - \exp(iy)| \leq |x - y|$, for x, y real,

$$\begin{aligned}|p(t - \tau_n(r', y')) - p(t - 2r'/c)| &= \left| \exp i\pi\gamma(t - \tau_n(r', y'))^2 - \exp i\pi\gamma(t - 2r'/c)^2 \right| \\ &= \left| \exp i\pi\gamma(v + (2r'/c - \tau_n(r', y')))^2 - \exp i\pi\gamma v^2 \right| \leq \pi\gamma \left| (v + (2r'/c - \tau_n(r', y')))^2 - v^2 \right| \\ &= \pi\gamma \left| (2v(2r'/c - \tau_n(r', y')) + (2r'/c - \tau_n(r', y'))^2) \right| \leq \pi\gamma (T\delta_1 + \delta_1^2)\end{aligned}$$

Next, the following is needed to bound the integral. Let $m = \min(\tau_n(r', y'), 2r'/c)$, $M = \max(\tau_n(r', y'), 2r'/c)$. Assume $\delta_1 \leq T$. Then

$$|\tau_n(r', y') - 2r'/c| \leq T \Leftrightarrow M - m \leq T \Leftrightarrow M - T/2 \leq m + T/2 \tag{2.14}$$

using the identity $\max(x, y) - \min(x, y) = |x - y|$. Further,

$$\begin{aligned}M - T/2 \leq t \leq m + T/2 &\Leftrightarrow M - t \leq T/2 \text{ and } t - m \leq T/2 \\ \Leftrightarrow \max(\tau_n(r', y') - t, 2r'/c - t) \leq T/2 &\text{ and } \min(t - \tau_n(r', y'), t - 2r'/c) \leq T/2 \\ \Rightarrow |t - \tau_n(r', y')| \leq T/2 \text{ and } |t - 2r'/c| \leq T/2\end{aligned}\tag{2.15}$$

and $m + T/2 - (M - T/2) = T - (M - m) = T - |\tau_n(r', y') - 2r'/c| \leq T$. Finally

$$\begin{aligned}
& \left| \frac{1}{T} \int (p(t - \tau_n(r', y')) - p(t - 2r'/c)) \bar{p}(t - 2r'/c) dt \right| \\
& \leq \frac{1}{T} \int \left| (p(t - \tau_n(r', y')) - p(t - 2r'/c)) \right| dt \\
& \leq \int_{m-T/2}^{M-T/2} \frac{1}{T} dt + \frac{1}{T} \int_{M-T/2}^{m+T/2} \pi \gamma (T\delta_1 + \delta_1^2) dt + \int_{m+T/2}^{M+T/2} \frac{1}{T} dt \leq 2\delta_1/T + \pi \gamma (T\delta_1 + \delta_1^2)
\end{aligned}$$

which completes the proof.

Lemma 2.2.

$$\begin{aligned}
& \left| \frac{1}{\#A(y)} \sum_{A(y)} \left(\exp(-i2\pi f_o \tau_n(r', y')) - \exp\left(-i2\pi f_o \frac{2}{c} \left(r' + \frac{1}{2} (y' - n\Delta a)^2 / r \right) \right) \right) \right. \\
& \quad \left. \exp\left(i2\pi f_o \frac{2}{c} \left(r + \frac{1}{2} (y - n\Delta a)^2 / r \right) \right) \right| \rightarrow 0
\end{aligned} \tag{2.16}$$

as $r \rightarrow \infty$ and $r' \rightarrow \infty$.

Proof.

$$\begin{aligned}
& \left| \exp(-i2\pi f_o \tau_n(r', y')) - \exp\left(-i2\pi f_o \frac{2}{c} \left(r' + \frac{1}{2} (y' - n\Delta a)^2 / r \right) \right) \right| \\
& \leq 2\pi f_o \left| \tau_n(r', y') - \frac{2}{c} \left(r' + \frac{1}{2} (y' - n\Delta a)^2 / r \right) \right| \\
& \leq 2\pi f_o \left(\left| \tau_n(r', y') - \frac{2}{c} \left(r' + \frac{1}{2} (y' - n\Delta a)^2 / r' \right) \right| + \frac{1}{c} \left| \frac{(y' - n\Delta a)^2}{r} - \frac{(y' - n\Delta a)^2}{r'} \right| \right) \\
& \leq 2\pi f_o \left(\delta_2 + \frac{1}{c} \sup_n \left((y' - n\Delta a)^2 / r + (y' - n\Delta a)^2 / r' \right) \right)
\end{aligned} \tag{2.17}$$

using (2.13), and the proof is evident.

Proof of Theorem 2.1. By (2.2), (2.7), and (2.11),

$$\begin{aligned}
k(r, y, r', y') &= \frac{1}{T} \int p(t - \tau_n(r', y')) \bar{p}(t - 2r/c) dt \\
& \quad \frac{1}{\#A(y)} \sum_{n \in A(y)} \exp(-i2\pi f_o \tau_n(r', y')) \exp\left(i2\pi f_o \frac{2}{c} \left(r + \frac{1}{2} (y - n\Delta a)^2 / r \right) \right)
\end{aligned} \tag{2.18}$$

It is a straightforward but tedious calculation which shows that

$$\begin{aligned}
& 1/T \int p(t - 2r'/c) \bar{p}(t - 2r/c) dt \\
&= 1/T (T - 2|r - r'|/c) \operatorname{sinc} \left((T - 2|r - r'|/c) \frac{2\gamma}{c} (r - r') \right) \mathbb{1}_{[-cT/2, cT/2]}(r - r') \\
&= g_{\gamma T^2} \left(\frac{r - r'}{c/2\gamma T} \right)
\end{aligned} \tag{2.19}$$

Hence Lemma 2.1 provides an error bound for approximating the function given by the integral term in (2.18) with $g_{\gamma T^2} \left(\frac{r - r'}{c/2\gamma T} \right)$. Further,

$$\begin{aligned}
& \frac{1}{\#A(y)} \sum_{n \in A(y)} \exp \left(-i2\pi f_0 \frac{2}{c} \left(r' + \frac{1}{2} (y' - n\Delta a)^2 / r \right) \right) \exp \left(i2\pi f_0 \frac{2}{c} \left(r + \frac{1}{2} (y - n\Delta a)^2 / r \right) \right) \\
&= \exp i2\pi \frac{2}{\lambda_0} (r - r') \frac{1}{\#A(y)} \sum_{n \in A(y)} \exp i2\pi \frac{1}{r\lambda_0} \left((y - n\Delta a)^2 - (y' - n\Delta a)^2 \right) \\
&\approx \exp i2\pi \frac{2}{\lambda_0} (r - r') \frac{1}{B} \int_{y-B/2}^{y+B/2} \exp i2\pi \frac{1}{r\lambda_0} \left((y - u)^2 - (y' - u)^2 \right) du \\
&= \exp i2\pi \frac{2}{\lambda_0} (r - r') \exp \left(-i2\pi \frac{1}{r\lambda_0} (y - y')^2 \right) \operatorname{sinc} \left(\frac{2B}{r\lambda_0} (y - y') \right) \\
&= g_{\infty} \left(\frac{y - y'}{r\lambda_0/2B} \right) \exp i\phi(r, y, r', y')
\end{aligned} \tag{2.20}$$

Thus, the function given by the sum term in (2.18) is estimated by

$$g_{\infty} \left(\frac{y - y'}{r\lambda_0/2B} \right) \exp i\phi(r, y, r', y'), \text{ which results from replacing } \exp(-i2\pi f_0 \tau_n(r', y')) \text{ with }$$

$\exp \left(-i2\pi f_0 \frac{2}{c} \left(r' + \frac{1}{2} (y' - n\Delta a)^2 / r \right) \right)$. The error bound for the sum in (2.18) is given by the sum of the bound of Lemma 2.2 with an error bound δ_4 for the Riemann sum estimation denoted by the “ \approx ” in the above (2.20).

The Riemann sum error bound δ_4 is given as follows. First, note that $(y-u)^2 - (y'-u)^2 = y^2 - y'^2 - 2(y-y')u$. Thus, $\exp i2\pi \frac{1}{r\lambda_o} (y^2 - y'^2)$ can be factored

out of both the sum and the integral, and it suffices to bound

$$\left| \frac{1}{\#A(y)} \sum_{n \in A(y)} f(n\Delta a) - \frac{1}{B} \int_{y-B/2}^{y+B/2} f(u) du \right|$$

where $f(u) = \exp i \frac{4\pi}{\lambda_o} \frac{(y-y')}{r} u$. Now,

$$\frac{1}{B} \int_{y-B/2}^{y+B/2} f(u) du = \frac{1}{B} \sum_{n \in A(y)} \int_{n\Delta a}^{(n+1)\Delta a} f(u) du = \frac{1}{B} \sum_{n \in A(y)} f(u_n) \Delta a$$

where the last equality follows from the mean value theorem, with $u_n \in [n\Delta a, (n+1)\Delta a]$.

Finally, using $\#A(y) = \frac{B}{\Delta a}$,

$$\left| \frac{1}{\#A(y)} \sum_{n \in A(y)} f(n\Delta a) - \frac{1}{B} \int_{y-B/2}^{y+B/2} f(u) du \right| \leq \frac{1}{\#A(y)} \sum_{n \in A(y)} |f(u_n) - f(n\Delta a)| \leq \frac{4\pi}{\lambda_o} \delta_3 \Delta a$$

using the bound

$$\begin{aligned} |f(u_n) - f(n\Delta a)| &= \left| \exp i \frac{4\pi}{\lambda_o} \frac{(y-y')}{r} u_n - \exp i \frac{4\pi}{\lambda_o} \frac{(y-y')}{r} n\Delta a \right| \\ &\leq \frac{4\pi}{\lambda_o} \frac{|y-y'|}{r} |u_n - n\Delta a| \leq \frac{4\pi}{\lambda_o} \delta_3 \Delta a \end{aligned}$$

where $\delta_3 = \frac{|y-y'|}{r}$.

The integral term and the sum term in (2.18) together with their estimating functions are all bounded by one. Therefore the product of the two terms is estimated by the product of their estimates, with a bound given by the sum of the bound of Lemma 2.1, that of Lemma 2.2, and δ_4 . This completes the proof, and also establishes bounds which can be used in practice to verify that approximations are valid.

Summary. It has now been established that

$$k(r, y, r', y') \approx g_{\gamma T^2} \left(\frac{r - r'}{c/2\gamma T} \right) g_{\infty} \left(\frac{y - y'}{r\lambda_o/2B} \right) \exp i\phi(r, y, r', y') \quad (2.21)$$

with

$$g_b(x) = (1 - |x|/b) \text{sinc}((1 - |x|/b)x) 1_{[-b, b]}(x) \quad (2.22)$$

Plots of $g_b(x)$ are given in Fig. 2.4. The first zero crossing of g_b is approximately equal to that of the sinc function for b large, i.e., $\inf\{x: g_b(x) = 0\} \approx \inf\{x: \text{sinc}(x) = 0\} = 1$, as $b \rightarrow \infty$, which is valid in (2.21) for a standard system in that γT^2 is sufficiently large. Since g_b is a real-valued function, the magnitude of k is given by the product of the two g_b functions in (2.21). The *range resolution* is approximately given as $c/\gamma T$, dictated by the first term; the *azimuth resolution*, which depends on range r , is $r\lambda_o/B$, dictated by the second term. Resolutions are defined as the width of the interval about zero on which g_b is positive.

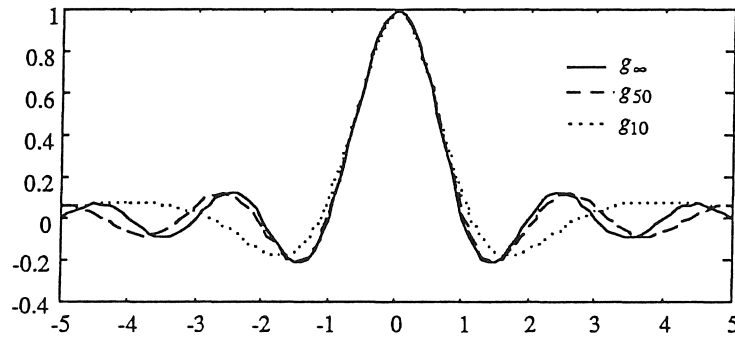


Fig. 2.4. g_b functions.

Surface Model

In this section a physical model is presented for the surface, which begins the establishment of the relationship between $a(r, y, \theta)$, the reflectance of the surface, and $\rho(y, \theta)$, the surface location function. The model is motivated as follows by the non-ideal resolution of the sensor. As mentioned previously, the radar image at (r, y) can be

thought of as the energy returned by the subset

$S_{(r,y)} = \{(r', y', \theta): r' \approx r, y' \approx y, (r', y', \theta) \in S\}$, where the accuracy associated with the “ \approx ” notation is precisely given by the resolution and functional form of the point spread function, as given above. In a typical practical setting this subset is comprised of a number of surface facets, randomly placed, representing fine-scale surface structure. The individual scatterers are beyond the resolving capability of the sensor system, and it is reasonable to express this by modeling small scale roughness, and, loosely speaking, interpreting a value $\rho(y, \theta)$ of the surface location function as providing the “average” position of all surface facets in a neighborhood of the location $(\rho(y, \theta), y, \theta)$. This discussion is made precise in the following.

Let L be a lattice, with associated grid $\{(y'_i, \theta_j): (i, j) \in L\}$, where $y'_i - y'_{i-1} = \Delta y$ and $\theta_j - \theta_{j-1} = \Delta \theta$. The azimuth spacing Δy and angular spacing $\Delta \theta$ are chosen small enough to allow representing the highest spatial frequency component appropriate given sensor resolution. Higher frequency components correspond to small-scale roughness. Figure 2.5 shows the sensing of a small patch of the surface S for some $(i, j) \in L$.

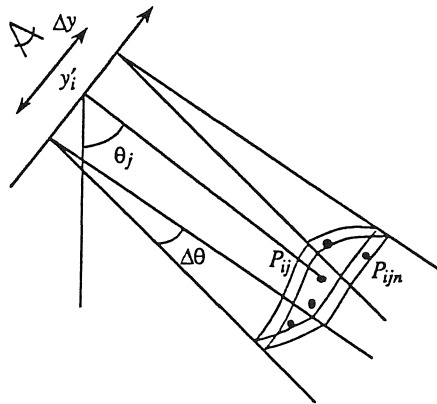


Fig. 2.5. Surface model.

A point $P_{ij} = (\rho(y'_i, \theta_j), y'_i, \theta_j)$ serves as the representative location for the patch, and the patch is assumed to consist of point scatterers $P_{ijn} = (\rho(y'_i, \theta_j) + \eta_{1ijn}, y'_i + \eta_{2ijn}, \theta_j + \eta_{3ijn})$

at random positions about the representative position, where there are N_{ij} scatterers, i.e., $n = 1, 2, \dots, N_{ij}$, and $\{(\eta_{1ijn}, \eta_{2ijn}, \eta_{3ijn}) : (i, j) \in L, n = 1, 2, \dots, N_{ij}\}$ are independent zero-mean random vectors, and for each $(i, j) \in L$, $\{(\eta_{1ijn}, \eta_{2ijn}, \eta_{3ijn}) : n = 1, 2, \dots, N_{ij}\}$ are identically distributed. Thus, there is some small-scale uncertainty in the position of the point scatterers, physically surface facets. Also, assume that for all (i, j, n) , $(\eta_{1ijn}, \eta_{2ijn}, \eta_{3ijn})$ has compact support, so scatterers are contained in a voxel about P_{ijn} . Each point P_{ijn} has a random complex reflectance a'_{ijn} independent of position (i.e., of $(\eta_{1ijn}, \eta_{2ijn}, \eta_{3ijn})$) and $\{a'_{ijn} : (i, j) \in L, n = 1, 2, \dots, N_{ij}\}$ are independent, and for each $(i, j) \in L$, $\{a'_{ijn} : n = 1, 2, \dots, N_{ij}\}$ are identically distributed. In particular, the complex reflectance function a is defined as $ra(r, y, \theta) = \sum_{(i,j) \in L} \sum_{n=1}^{N_{ij}} a'_{ijn} \delta_{P_{ijn}}(r, y, \theta)$, where $\delta_{P_{ijn}}$ is a Dirac delta function with mass at P_{ijn} .

Now, plugging into the image formation equation (2.6),

$$\begin{aligned}
Z(r, y) &= \iiint \sum_{(i,j) \in L} \sum_{n=1}^{N_{ij}} a'_{ijn} \delta_{P_{ijn}}(r', y', \theta) k(r, y, r', y') d\theta dr' dy' \\
&= \sum_{(i,j) \in L} \sum_{n=1}^{N_{ij}} a'_{ijn} k(r, y, \rho(y'_i, \theta_j) + \eta_{1ijn}, y'_i + \eta_{2ijn})
\end{aligned} \tag{2.23}$$

In the next section the distribution of Z is derived.

Speckle Statistics

In this section the probabilistic component of the image formation process is addressed; this results from small-scale surface roughness, and is usually referred to as *speckle noise*.

The following theorem establishes the distribution of $Z(r, y)$ for an ideal ($a_{ijn} = 1$) point scatterer. Let $\sigma_1 = c/2\gamma T$, $\sigma_2 = \lambda_o/2B$, $b_1 = \gamma T^2$, and let $r' = \rho(y'_i, \theta_j) + \eta_{1ijn}$, $y' = y'_i + \eta_{2ijn}$. Then by Theorem 2.1,

$$k(r, y, r', y') \approx g_{b_1} \left(\frac{r-r'}{\sigma_1} \right) g_{\infty} \left(\frac{y-y'}{r\sigma_2} \right) \exp i\phi(r, y, r', y') \quad (2.24)$$

is the response to a randomly placed point P_{ijn} , where $\phi(r, y, r', y')$ is given by (2.12). Note that the reciprocal of the carrier wavelength λ_o appears as a factor in the phase function ϕ in (2.12). As a result, a change in r' by $\lambda_o/2$ brings about a change in the complex exponential of one period. If η_{1ijn} has a sufficiently smooth distribution with standard deviation many times λ_o , i.e., $\phi(r, y, r', y')$ has a distribution which is wide and smooth relative to the interval $[0, 2\pi)$, then $\exp i\phi(r, y, r', y')$ has a distribution which is approximately the same as the distribution of $\exp i\phi'$, where ϕ' is uniformly distributed on $[0, 2\pi)$.² A typical value of λ_o is 10cm. A typical range resolution of 100m is clearly insufficient to clear up small-scale uncertainty even on the order of many times λ_o . Indeed, if the width of the distributions of η_{1ijn} and η_{2ijn} are narrow relative to the sensor resolution, then $|k(r, y, r', y')| \approx |k(r, y, \rho(y'_i, \theta_j), y'_i)|$ for all outcomes of $(\eta_{1ijn}, \eta_{2ijn})$. Hence the assumption is that the distribution of $(\eta_{1ijn}, \eta_{2ijn})$ is narrow relative to the resolution but wide relative to λ_o . This idea is presented in a slightly different context in Kelly *et al.* [7]. In this context, the precise statement is as follows.

Theorem 2.2. Let $\phi(r, y, r', y')$ have a continuous p.d.f. which is directly Riemann integrable in the sense of Feller [2, pp. 362-363] over the real line. Then

$g_{b_1} \left(\frac{r-r'}{\sigma_1} \right) g_{\infty} \left(\frac{y-y'}{r\sigma_2} \right) \exp i\phi(r, y, r', y')$ is approximately equivalent in distribution to $g_{b_1} \left(\frac{r-\rho(y'_i, \theta_j)}{\sigma_1} \right) g_{\infty} \left(\frac{y-y'_i}{r\sigma_2} \right) \exp i\phi'$, where $\phi' \sim U[0, 2\pi)$, in the following sense. As

$\sigma_1, \sigma_2 \rightarrow \infty$, with b_1 fixed,

²In engineering terminology, the phase response of the system is completely random.

$$\left| g_{b_1} \left(\frac{r-r'}{\sigma_1} \right) g_{\infty} \left(\frac{y-y'}{r\sigma_2} \right) - g_{b_1} \left(\frac{r-\rho(y'_i, \theta_j)}{\sigma_1} \right) g_{\infty} \left(\frac{y-y'_i}{r\sigma_2} \right) \right| \rightarrow 0$$

almost everywhere, and as $\lambda_o \rightarrow 0$, $\exp i\phi(r, y, r', y') \rightarrow \exp i\phi'$ in distribution.

Proof. Since $(\eta_{1ijn}, \eta_{2ijn})$ has compact support, $|r' - \rho(y'_i, \theta_j)|/\sigma_1 \rightarrow 0$ and $|y' - y'_i|/\sigma_2 \rightarrow 0$ almost everywhere, and $g_b \leq 1$. Therefore, to prove the first assertion it suffices to show that g_b is uniformly continuous, which is done in Lemma 2.3. The second assertion follows immediately from Lemma 2.4.

Lemma 2.3. For $b \in (0, \infty]$, g_b is uniformly continuous.

Proof. First note that the derivative of the sinc function is given by

$$\text{sinc}'(x) = \frac{1}{x}(\cos \pi x - \text{sinc } x) = \frac{1}{x}O(x^2) \quad (2.25)$$

so $c = \sup_{x \in \mathbf{R}} \text{sinc}'(x) < \infty$. The derivative of the sinc function is plotted in Fig. 2.6. By the mean value theorem, for $x, y \in \mathbf{R}$, $|\text{sinc}(x) - \text{sinc}(y)| \leq \sup_{z \in \mathbf{R}} |\text{sinc}'(z)| |x - y| = c|x - y|$.

Further, since $|(1 - |x|/b) - (1 - |y|/b)| \leq \frac{1}{b}|x - y|$, for $x, y \in [-b, b]$,

$$\begin{aligned} |(1 - |x|/b)x - (1 - |y|/b)y| &\leq |(1 - |x|/b)x - (1 - |x|/b)y| + |(1 - |x|/b)y - (1 - |x|/b)y| \\ &\leq |1 - |x|/b| |x - y| + \frac{|y|}{b} |x - y| \leq 2|x - y| \end{aligned} \quad (2.26)$$

so that $|\text{sinc}(1 - |x|/b)x - \text{sinc}(1 - |y|/b)y| \leq 2c|x - y|$. It is now evident that for $x, y \in \mathbf{R}$, with g_b given by (2.22),

$$|g_b(x) - g_b(y)| \leq |(1 - |x|/b)|2c|x - y| + \frac{1}{b}|x - y| \leq (2c + 1/b)|x - y| \quad (2.27)$$

and the proof is complete.

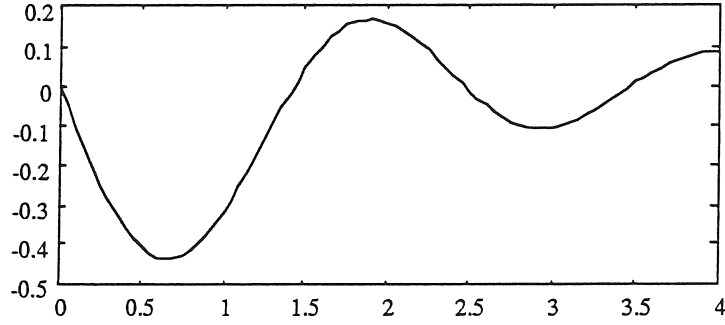


Fig. 2.6. Derivative of the sinc function.

Lemma 2.4. Let X be a random variable with continuous p.d.f. $f(x)$, directly R-integrable over \mathbf{R} . Then $\exp(i2\pi X/\gamma) \xrightarrow{d} \exp i\phi$ as $\gamma \rightarrow 0$, where $\phi \sim U[0, 2\pi)$.

Proof. $\frac{X}{\gamma} = \left[\frac{X}{\gamma} \right] + \Delta_\gamma$, where $\Delta_\gamma = \frac{X}{\gamma} - \left[\frac{X}{\gamma} \right]$, $\Delta_\gamma \in [0, 1)$. Further,

$\exp(i2\pi X/\gamma) = \exp i2\pi\Delta_\gamma$, so it suffices to show that $\Delta_\gamma \xrightarrow{d} U[0, 1)$. Let $t \in [0, 1)$.

$$P(\Delta_\gamma \leq t) = \sum_{n=-\infty}^{\infty} P(X \in [n\gamma, n\gamma + t\gamma)) = \sum_{n=-\infty}^{\infty} \int_{n\gamma}^{n\gamma+t\gamma} f(x) dx = \sum_{n=-\infty}^{\infty} f(u_n) t\gamma \quad (2.28)$$

where $u_n \in [n\gamma, n\gamma + t\gamma)$, and the last equality follows from the mean value theorem.

Finally, since f is R-integrable, $\sum_{n=-\infty}^{\infty} f(u_n) t\gamma \rightarrow t \int_{-\infty}^{\infty} f(u) du = t$, which completes the proof.

Incorporating the result of Theorem 2.2 in the image formation equation (2.23), the radar image is given by

$$\begin{aligned} Z(r, y) &\approx \sum_{(i,j) \in L} \sum_{n=1}^{N_{ij}} a'_{ijn} g_{b_1} \left(\frac{r - \rho(y'_i, \theta_j)}{\sigma_1} \right) g_\infty \left(\frac{y - y'_i}{r\sigma_2} \right) \exp i\phi'_{ijn} \\ &= \sum_{(i,j) \in L} g_{b_1} \left(\frac{r - \rho(y'_i, \theta_j)}{\sigma_1} \right) g_\infty \left(\frac{y - y'_i}{r\sigma_2} \right) C_{ij} \end{aligned} \quad (2.29)$$

where $C_{ij} = \sum_{n=1}^{N_{ij}} a'_{ijn} \exp i\phi'_{ijn}$, and $\{\phi'_{ijn} : (i, j) \in L, n = 1, 2, \dots, N_{ij}\}$ are i.i.d. $U[0, 2\pi)$,

independent of $\{a'_{ijn} : (i, j) \in L, n = 1, 2, \dots, N_{ij}\}$. Hence ϕ'_{ijn} represents small-scale

uncertainty in the location of facets, and a'_{ijn} represents the reflective properties of the facets.

A (zero-mean) complex normally distributed random variable Z has real and imaginary parts that are normal and i.i.d., and, as a consequence, $|Z|^2$ exponentially distributed, and $\arg Z$ uniformly distributed on $[0, 2\pi)$. Using the central limit theorem in the complex plane, applying it to the real and imaginary parts separately, as in Goodman [5], results in a complex normal distribution (implicitly taken to be zero mean) for

$$C_{ij} = \sum_{n=1}^{N_{ij}} a'_{ijn} \exp i\phi'_{ijn} \text{ (letting } N_{ij} \rightarrow \infty \text{)}. \text{ Since the sum of complex normal distributed}$$

random variables is complex normal, $Z(r, y)$ is also complex normal.

The value typically stored in a radar image is $|Z(r, y)|^2$, which due to $\arg Z(r, y)$ being uniform, is sufficient from a statistical perspective (relative to marginal distributions). Because $Z(r, y)$ is complex normal, $|Z(r, y)|^2$ has an exponential distribution. Since $\{C_{ij} : (i, j) \in L\}$ are zero mean and independent, by squaring (2.29) and taking the expectation,

$$\mathbb{E}|Z(r, y)|^2 = \sum_{(i, j) \in L} g_{b_1}^2 \left(\frac{r - \rho(y'_i, \theta_j)}{\sigma_1} \right) g_{\infty}^2 \left(\frac{y - y'_i}{r\sigma_2} \right) \mathbb{E}|C_{ij}|^2 \quad (2.30)$$

The quantity $\mathbb{E}|C_{ij}|^2$ has a standard physical interpretation as the expected returned energy from the surface element (voxel) represented by P_{ij} . Let $U(r, y) = |Z(r, y)|^2$. $\mathbb{E}U(r, y)$ represents the total expected returned energy from $S_{(r, y)}$. For (r, y) and (r', y') separated and k suitably impulsive $Z(r, y)$ and $Z(r', y')$ are approximately independent, in that off-diagonal elements of the covariance matrix are approximately zero. Hence, U is approximately independent exponential.

The basic result just established is that the distribution of $U(r, y)$ is approximately exponential with mean given by (2.30). This distribution is based on the assumption that

the surface is “rough relative to a wavelength” (see Goodman [5], the preceding makes this precise). Of course, there are many situations in which this assumption is not valid; there is plenty of room for other models. The claim here is simply that the model given above is appropriate for a large class of natural terrains. As support consider the following from Pettingill *et al.* [10]

At radar wavelengths the moon, Mercury, Venus and Mars appear to be densely covered by facets that vary in dimension from a few wavelengths to hundreds or even thousands of wavelengths that are tilted at random to the local horizontal. The same would be true of many points on the earth...

Multi-look SAR. An extension of the distributional form from exponential to the gamma family is often needed due to practical considerations. Consider the interval $[y - B/2, y + B/2)$ over which returned pulses are summed to obtain the radar image value at (r, y) . A *multi-look* SAR image is obtained by subdividing $[y - B/2, y + B/2)$ into disjoint intervals and for each element of the partition resolving what are independent realizations of the image based on the same scene. These i.i.d. exponential realizations are averaged, resulting in a gamma-distributed sum. The mean remains as above; however the variance decreases as $1/n$, where n is the number of *looks*, or independent realizations of the image. Since the azimuth resolution is inversely proportional to B , it degrades as the interval is subdivided. Thus resolution in azimuth can be traded off for a reduction in the speckle noise.

Note. Multi-look averaging is important for reduction of noise in practical situations. Let U have a gamma distribution, and let *signal-to-noise ratio* (SNR) be defined as $EU/\sqrt{\text{var}U}$ (called the *coefficient of variation* in statistics). For example, the one-look distribution has SNR 1, for four-look it is 2, both quite low. The averaging of looks is often referred to as *incoherent* averaging, since as in optics it refers to summing with respect to energy, i.e., the magnitudes squared of the complex-envelope radar image

values. The summing of complex envelopes is referred to as *coherent* averaging, again because of the analogy to optics.

Summary. There is a random component in the image formation process for the radarclinometry problem due to small-scale surface roughness beyond the resolution of the sensor. Surface location is described by a point set $\{P_{ij}:(i,j) \in L\}$ which represents a summary of local behavior. Random displacements of scatterers about these representatives gives rise to the energy for each site in the radar image being a random fluctuation about an expected energy. The energy is gamma distributed, with mean given by (2.30). The phase, or argument, of the complex-valued output of the SAR signal processing, not represented in the energy, is completely uninformative, as it is uniformly distributed, and hence independent of the surface shape or dielectric properties.

The randomization of phase brought about by the surface effectively transforms coherent source energy into incoherent returned energy. The final component of the image formation process is the relationship between the surface location function, in particular the point set $\{P_{ij} = (\rho(y'_i, \theta_j), y'_i, \theta_j):(i,j) \in L\}$, and $E|C_{ij}|^2$. In modeling the expected returned energy from the surface, because of the random phase, the appropriate physical perspective is that pertaining to incoherent energy, where interference effects are random. This model is developed in the next section.

Reflectance Function

A *reflectance function* accounts for variation in returned energy due to surface shape, and operates on the surface location function ρ . The radiometric law for the simple case in which the surface composition is homogeneous, often referred to as the *uniform albedo* assumption, and the reflectance function is Lambertian, is derived in this section using a standard technique of geometric optics (see, e.g., Boyd [1]). With Lambert reflectance, energy is radiated from the surface uniformly in all directions. This simple

model is extended later in the presentation. The next three paragraphs are preliminary to the development.

At this point the surface location function is defined only on the grid $\{(y'_i, \theta_j) : (i, j) \in L\}$. The technique of geometric optics used here requires that partial derivatives be defined. Extend the surface location function ρ to the real plane using the *low pass representation* (see Franks [4]):

$$\rho(y, \theta) = \sum_{(i, j) \in L} \text{sinc}((y - y'_i)/\Delta y) \text{sinc}((\theta - \theta_j)/\Delta \theta) \rho(y'_i, \theta_j) \quad (2.31)$$

i.e., the function is interpolated using a sinc-function kernel.

Note. By way of explaining the terminology, the low pass representation results from applying a spatial filter with ideal low pass transfer function $1_{[0, 1/2\Delta y] \times [0, 1/2\Delta \theta]}$ to the function $\sum_{(i, j) \in L} \rho(y'_i, \theta_j) \delta_{(y'_i, \theta_j)}$. Hence, the information contained in the samples $\{\rho(y'_i, \theta_j) : (i, j) \in L\}$ is represented but no high frequency components, about which the samples convey no information, are introduced.

In the following discussion the standard terminology of radiometry is used. Illumination power (energy per unit time) per unit area is called *irradiance*. The term is used to describe light energy incident on a surface. Light radiated from a surface is described by *radiance*, which is in units of power/(area · solid angle); solid angle is a measure of a set of directions, given by the area of the intersection of a set of rays pointing according to the set of directions, and the unit sphere.

Employ the simplifying assumption that for a point (r, y, θ) on the surface, the sensor is considered a point source of energy at $(0, y, \theta)$.

Note. The assumption is reasonable if the surface is distant from the flight path relative to B . This has already been assumed in deriving the approximate form of the point spread function.

Since the sensor is a point source, the power per unit solid angle radiated by the sensor is constant with respect to direction, i.e., energy propagates spherically.

With the preliminaries out of the way, let the surface be $S = \{(\rho(y, \theta), y, \theta)\}$. Fix (i, j) , and consider a patch of the surface $\delta S = \{(\rho(y, \theta), y, \theta) : |y - y_i| \leq \Delta y/2, |\theta - \theta_j| \leq \Delta \theta/2\}$, and let $r = \rho(y_i, \theta_j)$. Figure 2.7 depicts a cross section of the sensing of the patch. There is an angle α between the vector s pointing at the sensor and the normal n to the surface patch. The solid angle subtended by the patch is $\Delta y r \Delta \theta / r^2$, and the solid angle subtended by the sensor antenna is $\frac{\chi}{r^2}$, where χ is the area of the antenna (assume the distance from the antenna to the surface is large relative to the area of the antenna). Note that the notation does not express dependence on (i, j) of the quantities defined above for convenience.

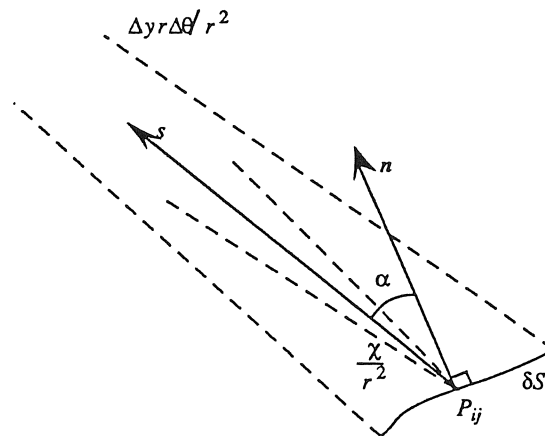


Fig. 2.7. Sensing of a surface patch.

The power returned to the sensor by the patch δS is calculated in order to obtain the relationship between returned energy and the surface shape. The integration of this

power with respect to time carried out by the SAR signal processing yields the energy of the return. Note that power and energy returned by the surface are in a strict sense *expected values*, but are not referred to as such in the following for ease of presentation.

The power incident on δS is given by $P \Delta y r \Delta \theta / r^2$ where P is the illumination power per unit solid angle radiated by the sensor (assumed a point source). To obtain the surface irradiance on the patch the incident power is divided by the area $\frac{\Delta y r \Delta \theta}{\cos \alpha}$ of δS , so the surface irradiance is $P \frac{\cos \alpha}{r^2}$.

For a Lambertian surface radiance is proportional to irradiance with proportionality constant $1/\pi$. Hence, the radiance of the patch is $\frac{1}{\pi} P \frac{\cos \alpha}{r^2}$. The power radiated in the direction of the sensor is calculated by multiplying the radiance by the apparent area $\Delta y r \Delta \theta$ of the patch and the solid angle subtended by the sensor; so finally, the power returned by the patch is given by $\frac{\chi}{\pi} P \frac{\cos \alpha}{r^4} \Delta y r \Delta \theta$.

Summary. For a surface patch represented by the point $(\rho(y'_i, \theta_j), y'_i, \theta_j)$, the expected energy returned is $E|C_{ij}|^2 = \frac{\cos \alpha(y'_i, \theta_j)}{\rho^3(y'_i, \theta_j)} \Delta y \Delta \theta$, where

$$\cos \alpha(y, \theta) = \frac{1}{\sqrt{(\rho_y(y, \theta))^2 + \left(\frac{\rho_\theta(y, \theta)}{\rho(y, \theta)}\right)^2 + 1}} \quad (2.32)$$

with subscripts denoting partial derivatives.

Thus, by (2.30), for a Lambertian surface of uniform albedo, the expected energy for the radar image is³

$$EU(r, y) \approx \text{const.} \sum_{(i,j) \in L} \kappa_r(r - \rho(y'_i, \theta_j), y - y'_i) \frac{\cos \alpha(y'_i, \theta_j)}{\rho^3(y'_i, \theta_j)} \Delta y \Delta \theta \quad (2.33)$$

where $\kappa_r(u, v) = g_{b_1}^2 \left(\frac{u}{\sigma_1} \right) g_{\infty}^2 \left(\frac{v}{r\sigma_2} \right)$. A mesh plot of κ_r for Magellan as used in one of the experiments presented later is shown in Fig. 2.8. Note the range resolution is worse than the azimuth resolution. The above (2.33) is a Riemann sum which approximates the integral

$$EU(r, y) \approx \text{const.} \iint \kappa_r(r - \rho(y', \theta), y - y') \frac{\cos \alpha(y', \theta)}{\rho^3(y', \theta)} dy d\theta \quad (2.34)$$

Before generalizing the radiometric law it is worth comparing the image formation process to that for standard shape-from-shading.

³Note that the point spread function is obtained by integrating over time; incorporating this with the constant $\frac{\lambda}{\pi} P$ implies that the SAR image value has units of energy, as expected.

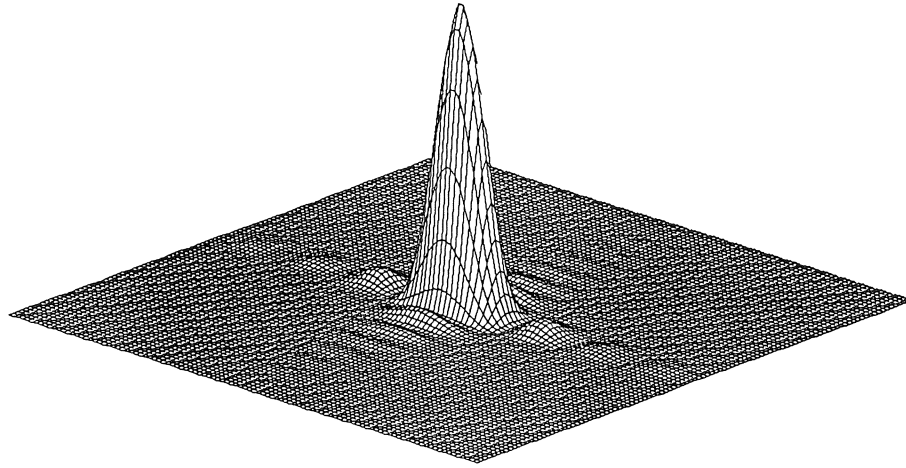


Fig. 2.8. Mesh plot of κ_r .

Comparison to Standard Shape-from-shading

The main goal here is to allow comparing the results derived above with the model in Frankot and Chellappa [3], in which radarclinometry is accomplished by shape-from-shading techniques. Shape-from-shading (Horn [6]), or photoclinometry, is the derivation of shape information from a standard image acquired with a frame camera under incoherent illumination (usually visible-spectrum). Procedures typically involve assumption of an ideal point spread function and parallel projection (described later).

The important distinction of radar imaging relative to standard frame sensing is that sensors employ different mechanisms for recording information about the illuminated surface. In standard shape-from-shading energy incident on an image plane is measured, and $\cos\alpha$ is expressed in terms of the image plane coordinate system, i.e., expressed as a function of the same variables used to index the image. The analogue here is to express $\cos\alpha$ as a function of (r,y) , *not* (y,θ) .

Toward this end, assuming it is well defined, let $\theta(r,y)$ take the value $\theta: r = \rho(y, \theta)$. Taking partial derivatives of the expression $r = \rho(y, \theta(r,y))$ with respect to r and y , using the chain rule,

$$\rho_\theta(y, \theta)|_{\theta=\theta(r,y)} = \frac{1}{\theta_r(r,y)}; \quad \rho_y(y, \theta)|_{\theta=\theta(r,y)} = \frac{-\theta_y(r,y)}{\theta_r(r,y)} \quad (2.35)$$

from which, using (2.32),

$$\cos\alpha(y, \theta(r,y)) = \frac{r\theta_r(r,y)}{\sqrt{(r\theta_r(r,y))^2 + (r\theta_y(r,y))^2 + 1}} \quad (2.36)$$

Of course, this may be obtained by computing $\cos\alpha$ directly relative to the function $\theta(r,y)$. Consistent with the usual assumption, let the point spread function be ideal in (2.34), i.e., let $\kappa_r(u, v) \equiv \delta_{(0,0)}(u, v)$, where δ denotes the Dirac delta function. After a change of variables $r' = \rho(y', \theta)$ and evaluation of the integral, the image formation equation (2.34) becomes

$$EU(r,y) \approx \text{const.} \frac{\cos\alpha(y, \theta(r,y))}{r^3} \theta_r(r,y) \quad (2.37)$$

Parallel Projection. Now also assume that *parallel projection* holds, defined as follows. See Fig. 2.9 below. There exists an r_o such that $\frac{r_o}{r} \approx 1$ and a θ_o such that $\sin(\theta(r,y) - \theta_o) \approx \theta(r,y) - \theta_o$, for all r and y . Define $u(r,y) = r_o(\theta(r,y) - \theta_o)$. Then

$$u(r,y) \approx r_o \sin(\theta(r,y) - \theta_o) \approx r \sin(\theta(r,y) - \theta_o) \quad (2.38)$$

which is the distance from a point located a distance r along the ray oriented at an angle θ_o , to the surface. In other words, there is a “direction” θ_o around which the angular spread of the surface is small. The term “parallel” refers to the fact that rays emanating from the source to the surface are approximately parallel.

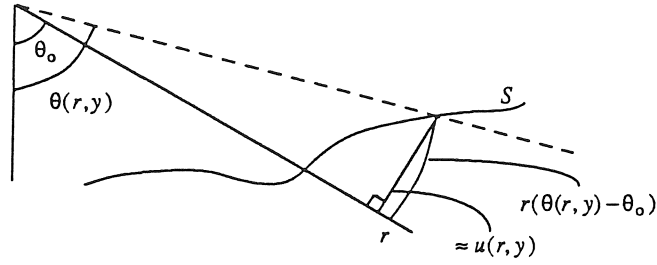


Fig. 2.9. Parallel projection.

Substituting into the image formation equation (2.37) above, using the fact that $\frac{1}{r^3} \approx \frac{1}{r_0^3}$,

$$EU(r,y) \approx \text{const.} \frac{u_r(r,y)}{\sqrt{(u_r(r,y))^2 + (u_y(r,y))^2 + 1}} u_r(r,y) \quad (2.39)$$

The first term in the product above is $\cos\alpha$ relative to a surface location function $u(r,y)$. Thus, under a Lambertian assumption the image is *not* proportional to $\cos\alpha$ expressed in the image coordinate system, which is the case in standard shape-from-shading for standard visible-light images. The reason for this is the geometry is different; there is no image plane on which power radiated in the direction of a normal to the (r,y) -plane $\{(r,y,\theta):\theta = \theta_0\}$ is measured. The power measured is that radiated *perpendicularly* to this direction, i.e., toward the flight path of the sensor. For a patch at (r,y,θ) , the power radiated in the direction of the flight path is proportional to $\cos\alpha \, dy \, du = \cos\alpha \, u_r(r,y) \, dr \, dy$, by a change of variables $u = u(r,y)$. The second term in the product above results from this change of variables. Returning to radarclinometry, the Lambert and uniform-albedo assumptions are relaxed after providing a couple of examples.

Examples

In this section examples of image formation are provided for two simple surfaces.

Flat Plate. Figure 2.10 depicts the geometry for the sensing of a flat plate. Assume an ideal point spread function, i.e., $\kappa_r = \delta_{(0,0)}$. Since $\cos\alpha(y, \theta(r, y)) = \cos\theta(r, y) = z_0/r$, from which $\theta_r(r, y) = 1/r\sqrt{(r/z_0)^2 - 1}$, the image formation equation (2.37) yields
$$EU(r, y) \approx \text{const.} \frac{z_0}{r^5 \sqrt{(r/z_0)^2 - 1}}.$$
 Note that the function $\theta(r, y)$ need not exist, as

illustrated in the next example.

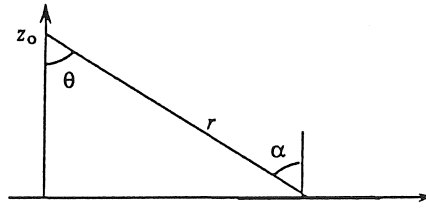


Fig. 2.10. Flat plate.

Cylindrical Section. Figure 2.11 shows the sensing of a cylindrical section. Assume $\kappa_r(u, v) = g_b^2(u)\delta_0(v)$. Here $\rho(y, \theta) = c$, and $\cos\alpha(y, \theta) = 1$. Hence, by image formation equation (2.34), $EU(r, y) \approx \text{const.} \frac{1}{c^3} g_b^2(r - c)$.

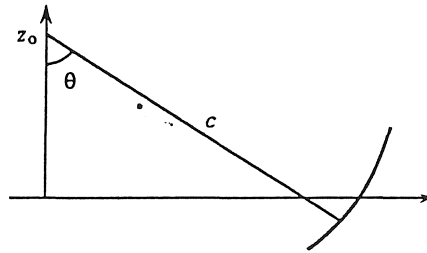


Fig. 2.11. Cylindrical section.

General Radiometric Model

In the general model presented here the expression for $E|C_{ij}|^2$ involves two elements: not only the reflectance function, which allows for variation in returned energy due to surface shape, but also the *albedo function* δ which describes reflective properties based on surface composition. The uniform albedo case corresponds to $\delta \equiv 1$. This second

element allows for variation in the dielectric properties of the surface. The model is given by

$$E|C_{ij}|^2 = \frac{f(\cos \alpha(y'_i, \theta_j))}{\rho^3(y'_i, \theta_j)} \delta(y'_i, \theta_j) \Delta y \Delta \theta \quad (2.40)$$

The cosine dependence is extended using the function f . The case where f is the identity function corresponds to the Lambert reflectance function. The case $f = (\bullet)^p$ for $p \geq 1$ is the Minnaert [8] or generalized Lambert function, which allows for the radiated distribution of energy to be concentrated orthogonally to surface patches, i.e., to be more *specular* than the Lambert model. The albedo function simply weights the expected energy. Note that a further generalization would have the map f depend on (y, θ) . In sufficiently restrictive situations this might be tenable; it is rejected simply for the sake of definiteness. The preceding results on image formation are summarized in the next section.

Image Formation

The radar image represents energy as a function of range and azimuth. At each site (r, y) is stored the energy returned by a portion of the surface, as determined by the point spread function that is specified by the SAR signal processing. The expected energy at (r, y) is given by

$$E_{\rho, \delta} U(r, y) \approx \text{const.} \iint \kappa_r(r - \rho(y', \theta), y - y') \frac{f(\cos \alpha(y', \theta))}{\rho^3(y', \theta)} \delta(y', \theta) dy' d\theta \quad (2.41)$$

where ρ is the surface location function, δ the albedo function, and $\alpha(y, \theta)$ is the angle between a vector normal to the surface at $(\rho(y, \theta), y, \theta)$ and a vector pointing at the sensor.

The distribution of the radar image is independent gamma.

References

1. R. W. Boyd, (1983). *Radiometry and the Detection of Optical Radiation*, New York: John Wiley and Sons.
2. W. Feller, (1966). *An Introduction to Probability Theory and its Applications*, Vol. II, New York: John Wiley and Sons.
3. R. T. Frankot and R. Chellappa, (1990). "Estimation of surface topography from SAR imagery using shape from shading techniques," *Artificial Intelligence*, 43, pp. 271-310.
4. L. E. Franks, (1981). *Signal Theory*, Stroudsburg, Pennsylvania: Dowden and Culver.
5. J. W. Goodman, (1975). "Statistical properties of laser speckle patterns," in *Laser Speckle and Related Phenomena*, edited by J. C. Dainty, Springer-Verlag.
6. B. Horn and M. Brooks, (1989). *Shape from Shading*, Cambridge: MIT Press.
7. P. A. Kelly, H. Derin and K. D. Hartt, (1988). "Adaptive segmentation of speckled images using a hierarchical random field model," *IEEE Trans. Acoust., Speech, Signal Processing*, 36, pp. 1628-1641.
8. M. Minnaert, (1941). "The reciprocity principle in lunar photometry," *Astrophys. J.*, 93, pp.403-410.
9. D. C. Munson, Jr., and R. L. Visentin, (1989). "A signal processing view of strip-mapping synthetic aperture radar," *IEEE Trans. Acoust., Speech, Signal Processing*, 37, pp. 2131-2147.
10. G. H. Pettingill, D. B. Campbell, and H. Masursky, (1980). "The surface of Venus," *Scientific American*, 243, pp. 54-65.

CHAPTER 3

RADARCLINOMETRY AS A STOCHASTIC INVERSE PROBLEM

In this chapter the radarclinometry problem is posed as a stochastic inverse problem. The approach is based on a Bayesian model, which has two components: 1) a model for encoding *a priori* constraints or knowledge about the surface; 2) a model for the *forward* process, i.e., the (random) transformation which takes the surface function to the observed data, the radar image. The second component was treated in the previous chapter. The stochastic *inverse* problem is to invert the forward process subject to the prior constraints, i.e., recover the surface from the radar image.

Previous work in Bayesian image analysis and stochastic regularization is vast. The reader is referred to Demoment [1] and Geman [2] for overviews.

Bayesian Formulation

The Bayesian formulation of radarclinometry is given as a special case of the following general model, so that comparisons can be made to other application areas in image analysis.

Stochastic Inverse Problems in Image Analysis. The situation is generically as follows. There are *data* $u = (u_s)_{s \in F}$ available, where F is an index set of *sites*. The data are viewed as an outcome of an *observation process* U , which is related to, i.e., statistically dependent on, an *attribute process* X . The goal is to recover the attributes from the data. For example, in radarclinometry the data is a radar image, and the attribute of interest is the surface shape which gave rise to the radar image. In image reconstruction the data is a degraded form of an attribute image, the undegraded picture. The attributes may be classification or boundary labels, as in remote sensing applications.

The attribute process represents some information of interest, related to, but not directly conveyed by, the data. Put this way the inverse problem is simply that of Bayesian estimation. However, there are further components which are specific to inverse problems in image analysis, given in the following.

Data Distribution. The observation process is related to the attribute process in the following special way. There is an *image formation operator* Φ which takes the attributes x to an ideal, or non-random, version of the data; in particular, $\Phi x = E(U|X = x)$. Further,

$$P(U \in du | X = x) = f_{\Phi x}(u) du \quad (3.1)$$

where f_{θ} is parametrized by its mean, i.e., $\int u f_{\theta}(u) du = \theta$, and has the form

$$-\log f_{\theta}(u) = \lambda \sum_{s \in F} (h(\theta_s, u_s) + g(u_s)) \quad (3.2)$$

where λ is positive, and the function h is the *penalty function*, with the property that $h \geq 0$ with equality only when its arguments are equal. Observe that (3.2) is a negative log-likelihood, so $\arg \min_x \sum_{s \in F} h((\Phi x)_s, u_s)$ is the maximum likelihood estimate of x . The

function h penalizes deviations of the data u from the synthetic rendering $\Phi x = E(U|X = x)$ of the data based on x . Note that (3.1) and (3.2) are equivalent to $\{U_s\}_{s \in F}$ being independent conditioned on $X = x$. Also, the conditional distribution of the observation process, determined by the physics of image formation, depends on x only through Φ . Examples of distributions arising in image analysis which have the form (3.1), (3.2) are given next.

Examples. Consider first the penalty function in radarclinometry. An independent gamma distribution with mean θ has p.d.f.

$$f_{\theta}(u) = \prod_{s \in F} \frac{1}{\Gamma(\lambda)} \left(\frac{\lambda}{\theta_s} \right)^{\lambda} u_s^{\lambda-1} \exp\left(-\frac{\lambda}{\theta_s} u_s \right) \quad (3.3)$$

so $h(a,b) = b/a - \log(b/a) - 1$, and λ in (3.2) as well as in (3.3) is the number of looks associated with the SAR signal processing (see the section on multi-look SAR in Chapter 2). A plot of $h(1, \bullet)$ for radarclinometry is shown in Fig. 3.1. For the Poisson distribution, which arises in tomography and astronomy

$$h(a,b) = \begin{cases} b(a/b - \log(a/b) - 1) & b \geq 1 \\ a & b = 0 \end{cases}$$

For additive independent Gaussian noise, which arises in image segmentation and reconstruction, $h(a,b) = (a - b)^2$.

In the Gaussian case the penalty is based on a difference. For the gamma distribution the penalty is based on the *ratio* b/a . Gamma noise is often termed “multiplicative” because a gamma-distributed random variable has standard deviation (noise) proportional to the mean. The posterior distribution for the Bayesian formulation of image analysis problems is given in the next paragraph.

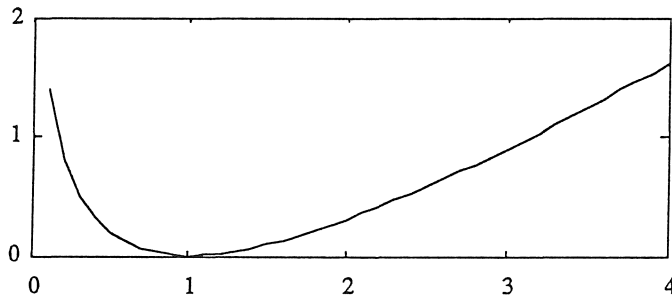


Fig. 3.1. Penalty function.

Posterior Distribution. Assume that the prior distribution has the Gibbs form $P(X \in dx) = \text{const.} \exp(-H(x)) dx$. By Bayes rule, the posterior distribution is

$$P(X \in dx | U = u) = \text{const.} \exp\left(-H(x) - \lambda \sum_{s \in F} h((\Phi(x))_s, u_s)\right) dx \quad (3.4)$$

Note that the normalizing constant depends on u . The Bayes point estimates used in image analysis are the mean of, the mode of, or a sample from the posterior distribution. The posterior also has the Gibbs form, with *energy*

$$H(x) + \lambda \sum_{s \in F} h((\Phi x)_s, u_s) \quad (3.5)$$

The first term is the prior, or *regularization term*, the second the *data term*. Note that the energy (3.5) could simply be regarded as a cost function, avoiding the Bayesian perspective completely. This is the approach in *regularization* (Poggio *et al.* [3]). Minimizing the energy (cost) yields the mode of the posterior distribution, the maximum *a posteriori* (MAP) estimate.

In this presentation on radarclinometry the perspective of Bayesian estimation is taken because, as is not always the case for other standard problems in image analysis, associating the *a priori* constraints with a prior distribution makes sense, for the following reason: realizations of the prior distribution assumed are realistic, in that, with high probability, they have the characteristics expected of the surface. The surface location function is essentially a height function, and samples from the prior distribution (described in the next chapter) placed on the space of surface functions are believable natural terrains, i.e., planetary surfaces. This is appropriate for the remote sensing applications of interest here. As a final point, note that for radarclinometry since λ is the number of looks, with which speckle noise variance decreases, it is intuitively satisfying that in (3.5) the data term is emphasized relative to the prior term as the number of looks increases.

Summary. Many image analysis problems fit within the framework used in posing radarclinometry as a stochastic inverse problem. What distinguishes image analysis problems from general Bayesian estimation is that statistical dependence between the data and the attributes of interest occurs through an image formation operator. Another

important consideration, which will be addressed later, is that the typically large data dimensionality requires design of efficient sampling algorithms used to obtain point estimates.

Discrete Image Formation Model

In this section, a discrete model (suitable for digital computation) is developed based on the material of Chapter 2. It is assumed that albedo is uniform and the reflectance function is Lambert. Some definitions are given first.

The radar image is indexed by $(r, y) \in F \subset [0, \infty) \times (-\infty, \infty)$. A reference surface ρ_o is chosen which is a good initial guess for the unknown surface function. Precisely, the surface process X satisfies $EX = 0$, where an outcome x is given by $\rho(y'_i, \theta_j) = \rho_o(y'_i, \theta_j) + x_{ij}$, so x_{ij} is the deviation from the reference surface. The prior distribution is placed on these deviations. It is reasonable to base the reference surface on a reference *geoid*, an analytical model of the shape of a planet. For example, the geometry for a “flat-earth” model is given in Fig. 3.2, where the reference surface is based on the distance from the flight path to grid points which are contained in a plane that approximates sea level locally.

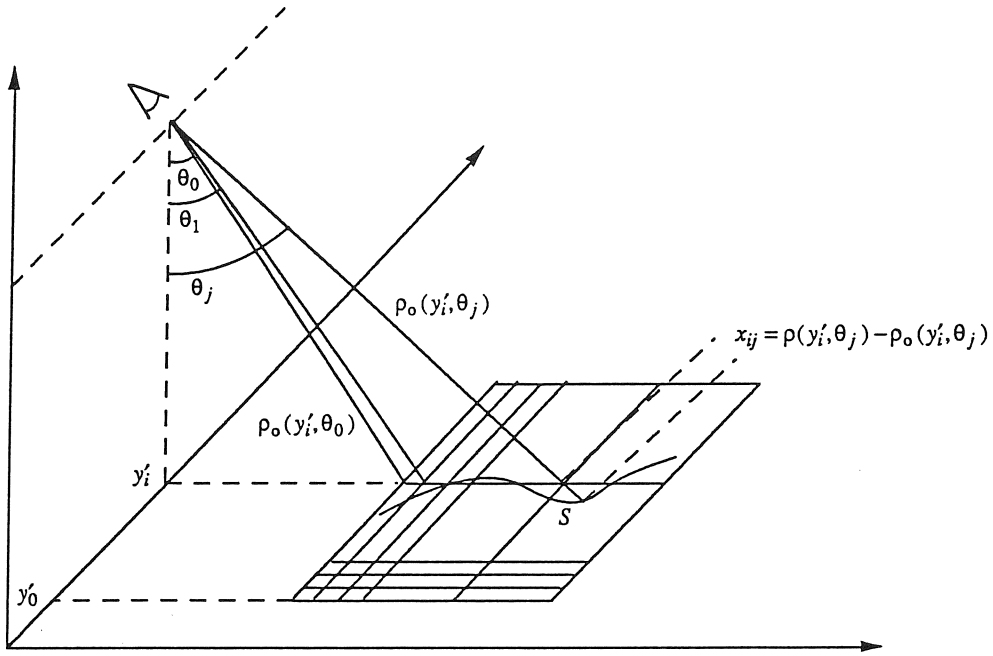


Fig. 3.2. Reference surface for flat earth model.

Let $(\Phi x)_s = \gamma \varphi_s(x)$; hence φ_s provides the image formation model up to a multiplicative constant, which accounts for gain due to unknown uniform albedo, antenna response, atmospheric attenuation, etc. The discrete image formation equation is given by (2.33), with partial derivatives approximated by differences:

$$\varphi_{(r,y)}(x) = \sum_{(i,j) \in L} \kappa_r \left(r - (x_{ij} + \rho_0(y_i', \theta_j)), y - y_i' \right) \frac{\cos \alpha_{ij}}{(x_{ij} + \rho_0(y_i', \theta_j))^3} \Delta y \Delta \theta \quad (3.6)$$

where

$$\cos \alpha_{ij} = \left(\left(\frac{\rho(y_i', \theta_j) - \rho(y_{i-1}', \theta_j)}{\Delta y} \right)^2 + \left(\frac{\rho(y_i', \theta_j) - \rho(y_i', \theta_{j-1})}{\rho(y_i', \theta_j) \Delta \theta} \right)^2 + 1 \right)^{-1/2} \quad (3.7)$$

The remaining component of the model is the prior distribution for X , the surface process, given as follows. The regularization energy is in the form of a Tikhonov stabilizer, frequently used in regularization problems. The next chapter presents

properties of the corresponding Gibbs distribution. There is irony in the fact that using a regularizer as the energy in a Gibbs distribution results in statistical characteristics that support the Bayesian perspective. In other words, samples from the resulting distributions are believable natural terrains.

References

1. G. Demoment, (1989). "Image reconstruction and restoration: overview of common estimation structures and problems," *IEEE Trans. Ac. Speech and Sig. Proc.*, 37, pp. 2024-2036.
2. D. Geman, (1991). "Random fields and inverse problems in imaging," *Lecture Notes in Mathematics*, Vol. 1427, Springer-Verlag.
3. T. Poggio, V. Torre, and C. Koch, (1989). "Computational vision and regularization theory," *Nature*, 317, pp. 314-319.

CHAPTER 4

REPRESENTATION OF STABILIZING PRIORS

This chapter provides results on the representation of a random field obtained by using a Tikhonov stabilizer (Tikhonov and Arsenin [9]) as the energy in a Gibbs distribution. The resulting distribution is often used as a prior in image analysis and computational vision (Poggio *et al.* [7]), and is also related to “intrinsic” fields (Künsch [4]).

The representation given here provides a method for generating samples (without using stochastic relaxation), allows deriving the maximum likelihood estimator for a scaling parameter, and gives a method for computing covariance structure. A self-similarity property is described.

The organization of the presentation is as follows. First, the distribution is defined in a general setting, and then a representation theorem is proved, based on which useful properties are derived. A special form of the distribution which is frequently used in practice is then assumed for the remainder of the presentation, for which other properties such as self-similarity are valid.

Stabilizing Gibbs Distribution

Joint Distribution. Consider an image x with domain $L = \{0, 1, \dots, n_1 - 1\} \times \{0, 1, \dots, n_2 - 1\}$. Let t_1, t_2 denote first-difference operators:

$$\begin{aligned}(t_1 x)(i_1, i_2) &= x(i_1, i_2) - x(i_1 - 1, i_2) \\ (t_2 x)(i_1, i_2) &= x(i_1, i_2) - x(i_1, i_2 - 1)\end{aligned}\tag{4.1}$$

The boundaries are accommodated by periodic extension; i.e., an index $(i_1, i_2) \notin L$ is taken to be $(i_1 \bmod n_1, i_2 \bmod n_2)$. This issue will be taken up again later. The operators t_1, t_2

may be thought of as discrete partial derivatives. This chapter concerns a random field X with outcomes in $\mathbf{R}^{n_1 n_2}$ and Gibbs distribution

$$P(X \in dx_{(0)}) = \text{const.} \exp\left(-\frac{1}{2} H(x_{(0)})\right) dx_{[0]} \quad (4.2)$$

with

$$H(x) = \sum_{(i_1, i_2) \in L} \sum_{1 \leq k+l \leq p} a_{kl} \left((t_1^k t_2^l x)(i_1, i_2) \right)^2 \quad (4.3)$$

where $_{(0)}$ denotes that the $(0,0)$ element is zero, and $_{[0]}$ that it is left out; i.e., $x_{(0)}(0,0) = 0$ and $x_{[0]} = (x_{(0)}(i_1, i_2))_{(i_1, i_2) \in L \setminus (0,0)}$. As given above H is the Tikhonov stabilizer of order p on the two-dimensional lattice, which is quadratic, so the distribution is Gaussian. The constraint $X(0,0) = 0$ is necessary for the distribution to exist. With this condition and each constant $a_{kl} \geq 0$ and $a_{0p}, a_{p0} > 0$, the normalizing constant is finite, and is calculated later. It is useful to express H as $H(x) = \|Tx\|^2$, where

$$Tx = (\sqrt{a_{10}} t_1 x, \sqrt{a_{01}} t_2 x, \sqrt{a_{20}} t_1^2 x, \sqrt{a_{11}} t_1 t_2 x, \sqrt{a_{02}} t_2^2 x, \dots, \sqrt{a_{0p}} t_2^p x) \quad (4.4)$$

a vector of all weighted partial derivative operators up to order p . Let $m = (p+1)(p+2)/2 - 1$, the number of elements in T . For $j = 1, 2, \dots, m$, T_j is used to denote the j^{th} element of T . In view of the fact that T is invariant to an additive constant, it is beneficial to view an outcome $x_{(0)}$ as the representative of an equivalence class $\{x_{(0)} + c; c \in \mathbf{R}\}$. The distribution is referred to here as the *stabilizing Gibbs distribution of order p* .

Local Distributions. Let $G: \{z \in \mathbf{C}: |z|=1\}^2 \rightarrow \mathbf{C}^m$ be a vector of polynomials given by

$$G(z_1, z_2) = (\sqrt{a_{10}}(1-z_1), \sqrt{a_{01}}(1-z_2), \sqrt{a_{20}}(1-z_1)^2, \sqrt{a_{11}}(1-z_1)(1-z_2), \sqrt{a_{02}}(1-z_2)^2, \dots, \sqrt{a_{0p}}(1-z_2)^p) \quad (4.5)$$

The domain of G is restricted to unit circles so that $\bar{z}_1 = z_1^{-1}$, $\bar{z}_2 = z_2^{-1}$, and $\|G\|^2$ can be expressed as a polynomial in z_1 and z_2 . Direct calculation by completing the square in the quadratic Gibbs energy $H(x)$ yields

$$\text{var}(X(i_1, i_2) | X(j_1, j_2) = x(j_1, j_2), (j_1, j_2) \neq (i_1, i_2)) = \frac{1}{b_{(0,0)}} \quad (4.6)$$

$$E(X(i_1, i_2) | X(j_1, j_2) = x(j_1, j_2), (j_1, j_2) \neq (i_1, i_2)) = \sum_{(j_1, j_2) \neq (i_1, i_2)} \frac{-b_{(j_1, j_2)}}{b_{(0,0)}} x(i_1 + j_1, i_2 + j_2) \quad (4.7)$$

where $b_{(j_1, j_2)}$ is the coefficient of $z_1^{j_1} z_2^{j_2}$ in the polynomial $\|G\|^2$. The coefficients $b_{(j_1, j_2)} = 0$ except for (j_1, j_2) in a neighborhood of the origin, so the field is Markov with a relatively local neighborhood structure. The local distributions given by the local variance (4.6) and local mean (4.7) represent conditional autoregressions (Besag [2], Ripley [8]), in particular, *intrinsic* autoregressions (Künsch [4]) because of the form of the coefficients.

Example. For the first order model $Tx = (t_1x, t_2x)$,

$$H(x) = \|Tx\|^2 = \sum_{(i_1, i_2) \in L} (x(i_1, i_2) - x(i_1 - 1, i_2))^2 + (x(i_1, i_2) - x(i_1, i_2 - 1))^2$$

$G(z_1, z_2) = ((1 - z_1), (1 - z_2))$, $\|G(z_1, z_2)\|^2 = 4 - z_1 - z_1^{-1} - z_2 - z_2^{-1}$. Hence, the local variance is 1/4, the local mean is

$$\frac{1}{4}(x(i_1 - 1, i_2) + x(i_1 + 1, i_2) + x(i_1, i_2 - 1) + x(i_1, i_2 + 1))$$

which involves only the four nearest neighbors of (i_1, i_2) . The following definitions are needed later.

Preliminaries

Complex Normal Distribution. A random vector $Z \in \mathbb{C}^n$ has a complex normal distribution $CN(\Sigma)$ if

$$P(Z \in dz) = \frac{1}{\pi^n \det(\Sigma)} \exp(-z' \Sigma^{-1} z) dz \quad (4.8)$$

where $\Sigma = E(ZZ')$, a prime denotes transposition and complex conjugation, and dz denotes $dx dy$ with $x = \text{Re}(z)$ and $y = \text{Im}(z)$. See Goodman [3] for standard properties.

For a function (matrix) ζ taking L to \mathbb{C} , let there be an understood ordering on L , so that ζ can be considered a vector, also denoted $(\zeta_{(i_1, i_2)})_{(i_1, i_2) \in L}$.

Discrete Fourier Transform. Let $z: L \rightarrow \mathbb{C}$, and $w_1 = \exp(\frac{i2\pi}{n_1})$, $w_2 = \exp(\frac{i2\pi}{n_2})$.

The discrete Fourier transform \hat{z} is the representation of z with respect to the orthonormal Fourier basis $\{e_{(k_1, k_2)}: (k_1, k_2) \in L\}$, where $e_{(k_1, k_2)} = ((n_1 n_2)^{-1/2} w_1^{i_1 k_1} w_2^{i_2 k_2})_{(i_1, i_2) \in L}$. In other

words, $z = P\hat{z}$ for a unitary matrix P whose columns consist of the elements of the Fourier basis, and $\hat{z} = P'z$. Further, $\|z\|^2 = \|P\hat{z}\|^2 = \hat{z}'P'P\hat{z} = \|\hat{z}\|^2 = \|P'z\|^2$, which is

Parseval's relation. The discrete Fourier transform of a multi-dimensional image $(z_i: L \rightarrow \mathbb{C})_{i=1,2,\dots,n}$ is defined as $\text{dft}((z_i)_{i=1,2,\dots,n}) = (\hat{z}_i)_{i=1,2,\dots,n}$. The main result of the

chapter is presented next.

Representation

Let

$$\Lambda(k_1, k_2) = (\sqrt{a_{10}}(1 - w_1^{-k_1}), \sqrt{a_{01}}(1 - w_2^{-k_2}), \sqrt{a_{20}}(1 - w_1^{-k_1})^2, \sqrt{a_{11}}(1 - w_1^{-k_1})(1 - w_2^{-k_2}), \sqrt{a_{02}}(1 - w_2^{-k_2})^2, \dots, \sqrt{a_{0p}}(1 - w_2^{-k_2})^p) \quad (4.9)$$

It is easy to verify that T_j has the eigenvalue $\Lambda_j(k_1, k_2)$ relative to the Fourier eigenvector $e_{(k_1, k_2)}$ for $(k_1, k_2) \in L$ and all j , where Λ_j denotes the j^{th} element of Λ . Note

that the $\{T_j; j = 1, 2, \dots, m\}$ have common eigenvectors, so in a sense $\Lambda(k_1, k_2)$ is a vector-valued eigenvalue of T . Now

$$P'T_j z = P'T_j P\hat{z} = P'P \operatorname{diag}\left(\Lambda_j(k_1, k_2)\right)_{(k_1, k_2) \in L} P'P\hat{z} = \operatorname{diag}\left(\Lambda_j(k_1, k_2)\right)_{(k_1, k_2) \in L} \hat{z} \quad (4.10)$$

so that for $(k_1, k_2) \in L$,

$$\left(P'T_j z\right)(k_1, k_2) = \Lambda_j(k_1, k_2) \hat{z}(k_1, k_2) \quad (4.11)$$

This states that $(\operatorname{dft}(Tz))(k_1, k_2) = \Lambda(k_1, k_2) \hat{z}(k_1, k_2)$, which can be verified directly. Also, denote $L \setminus (0, 0)$ by $L_{\{0\}}$.

Theorem 4.1. Let $(U(k_1, k_2))_{(k_1, k_2) \in L_{\{0\}}} \sim \operatorname{CN}(2I)$, where I is the identity matrix, so that U is i.i.d. complex normal. For $(i_1, i_2) \in L$, let

$$V(i_1, i_2) = \sum_{(k_1, k_2) \in L_{\{0\}}} \frac{U(k_1, k_2)}{\|\Lambda(k_1, k_2)\|} \left(e_{(k_1, k_2)}(i_1, i_2) - e_{(0, 0)}(i_1, i_2) \right) \quad (4.12)$$

Then $\operatorname{Re}(V)$ has a stabilizing Gibbs distribution of order p .

Note. This representation provides a method of generating a sample from the distribution based on an i.i.d. normal sample.

Proof. Follows immediately from Lemmas 4.1 and 4.2.

Lemma 4.1. Let $W \sim \operatorname{CN}(2I)$, with outcomes in $\mathbb{C}^{m n_1 n_2}$, so that W_j has outcomes in $\mathbb{C}^{n_1 n_2}$ for $j = 1, 2, \dots, m$. Let

$$Z = \operatorname{argmin}_{z_{(0)}: L \rightarrow \mathbb{C}} \|Tz_{(0)} - W\|^2 \quad (4.13)$$

where again, Z may be considered to have outcomes which are equivalence classes. Then $X = \operatorname{Re}(Z)$ has a stabilizing Gibbs distribution of order p .

The proof is an application of Lemma 4.3, given later.

Note. As an intuitive motivation, note that Z is the projection of “complex white noise” into the subspace $\{Tz_{(0)}: z_{(0)} \in \mathbb{C}^{n_1 n_2 - 1}\}$ of m -dimensional images Tz consisting of weighted discrete partial derivatives of z .

Example. Again, let $Tx = (t_1 x, t_2 x)$. Then

$$Z = \operatorname{argmin}_{z_{(0)}: L \rightarrow \mathbb{C}} \left\| \begin{bmatrix} t_1 z_{(0)} \\ t_2 z_{(0)} \end{bmatrix} - \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \right\|^2$$

Thus, Z is the projection of white noise onto the subspace of image pairs which are discrete partial derivatives. In other words, the partials of Z best match the i.i.d. variables W .

Lemma 4.2. V and Z have the same distribution.

Proof. It is convenient to use a Fourier basis in representing the above projection (4.13).

By Parseval’s relation and (4.11),

$$\begin{aligned} \|Tz - W\|^2 &= \sum_{j=1}^m \|T_j z - W_j\|^2 = \sum_{j=1}^m \|P' T_j z - P' W_j\|^2 \\ &= \sum_{(k_1, k_2) \in L} \sum_{j=1}^m \left| (P' T_j z)(k_1, k_2) - (P' W_j)(k_1, k_2) \right|^2 \quad (4.14) \\ &= \sum_{(k_1, k_2) \in L} \sum_{j=1}^m \left| \Lambda_j(k_1, k_2) \hat{z}(k_1, k_2) - (P' W_j)(k_1, k_2) \right|^2 \end{aligned}$$

Hence, the dft of Z is given by

$$\hat{Z} = \operatorname{argmin}_{\hat{z}: L \rightarrow \mathbb{C}, z(0,0)=0} \sum_{(k_1, k_2) \in L} \left\| \Lambda(k_1, k_2) \hat{z}(k_1, k_2) - \hat{W}(k_1, k_2) \right\|^2 \quad (4.15)$$

Because the Fourier basis is orthonormal, the dft of complex normal white noise is complex normal white noise with the same variance (this is the reason for embedding the projection in a complex space). The minimization may be carried out separately for each

$(k_1, k_2) \in L$. Note that $\|\Lambda(0,0)\| = 0$ so that $\hat{Z}(0,0)$ would be arbitrary were it not for the constraint $Z(0,0) = 0$, which requires that $\hat{Z}(0,0) = - \sum_{(k_1, k_2) \in L_{\{0\}}} \hat{Z}(k_1, k_2)$.

For $(k_1, k_2) \in L_{\{0\}}$, $\|\Lambda(k_1, k_2)\| > 0$, the $\hat{Z}(k_1, k_2)$ are independent, and are given by

$$\hat{Z}(k_1, k_2) = \frac{\Lambda(k_1, k_2)' \hat{W}(k_1, k_2)}{\|\Lambda(k_1, k_2)\|^2} \stackrel{d}{=} \frac{U(k_1, k_2)}{\|\Lambda(k_1, k_2)\|}$$

With $U(0,0) = - \sum_{(k_1, k_2) \in L_{\{0\}}} U(k_1, k_2)$, $(U(k_1, k_2))_{(k_1, k_2) \in L} \stackrel{d}{=} (\hat{Z}(k_1, k_2))_{(k_1, k_2) \in L}$, and applying

the inverse dft yields the result.

Note. The only dependence among the Fourier random variables $\hat{Z}(k_1, k_2)$ results from the constraint $\sum_{(k_1, k_2) \in L} \hat{Z}(k_1, k_2) = 0$. This is not true for \hat{X} since there are further constraints

to ensure that X is real.

Lemma 4.3. Let $W \sim \text{CN}(2I)$, with outcomes in \mathbb{C}^n , and let a real-valued $n \times p$ matrix A have full rank. Let

$$Z = \arg \min_{z \in \mathbb{C}^p} \|Az - W\|^2 \quad (4.16)$$

Then $P(X \in dx) = \text{const.} \exp(-\frac{1}{2} \|Ax\|^2) dx$, where $X = \text{Re}(Z)$.

Proof. Because A is real,

$$\begin{aligned} \|Az - W\|^2 &= \|\text{Re}(Az - W)\|^2 + \|\text{Im}(Az - W)\|^2 \\ &= \|A \text{Re}(z) - \text{Re}(W)\|^2 + \|A \text{Im}(z) - \text{Im}(W)\|^2 \end{aligned} \quad (4.17)$$

Hence the minimization of (4.16) can be carried out separately for the real and imaginary parts, and

$$X = \text{Re}(Z) = \arg \min_{x \in \mathbb{R}^p} \|Ax - \text{Re}(W)\|^2 \quad (4.18)$$

Since $\text{Re}(W) \sim N(0, I)$ and A has full rank, the minimization is a standard least-squares problem, so $X \sim N(0, (A'A)^{-1})$. The result is now obvious using $\|Ax\|^2 = x'A'Ax$.

Proof of Lemma 4.1. In Lemma 4.3, let A satisfy $Az_{[0]} = Tx_{(0)}$, so that mn_1n_2 plays the role of n in Lemma 4.3, and $n_1n_2 - 1$ the role of p . The matrix A so defined has full rank because $\det(A'A)$ exists and is non-zero, as shown in the next paragraph. The result is now evident, as $\|Tx_{(0)}\|^2 = \|Ax_{[0]}\|^2$.

Properties

Normalizing Constant. The normalizing constant for the stabilizing Gibbs distribution is computed as follows. Let $Ax_{[0]} = Tx_{(0)}$ and $X = \text{Re}(Z)$, as before. Since $X_{[0]}$ is normally distributed, the normalizing constant is given by $((2\pi)^{(n_1n_2-1)} \det((A'A)^{-1}))^{-1/2}$. To calculate $\det(A'A)$, first note that

$$\begin{aligned} \|Tx\|^2 &= \sum_{j=1}^m \|T_j x\|^2 = \sum_{j=1}^m \|P'T_j x\|^2 = \sum_{j=1}^m \sum_{(k_1, k_2) \in L} |\Lambda_j(k_1, k_2) \hat{x}(k_1, k_2)|^2 \\ &= \sum_{(k_1, k_2) \in L} |\hat{x}(k_1, k_2)|^2 \sum_{j=1}^m |\Lambda_j(k_1, k_2)|^2 = \sum_{(k_1, k_2) \in L} \|\Lambda(k_1, k_2)\|^2 |\hat{x}(k_1, k_2)|^2 \end{aligned} \quad (4.19)$$

so that

$$\begin{aligned} x'_{[0]} A'A x_{[0]} &= \|Tx_{(0)}\|^2 = \sum_{(k_1, k_2) \in L} \|\Lambda(k_1, k_2)\|^2 |(P'x_{(0)})(k_1, k_2)|^2 \\ &= x'_{(0)} P \text{diag} \left(\|\Lambda(k_1, k_2)\|^2 \right)_{(k_1, k_2) \in L} P' x_{(0)} \\ &= x'_{[0]} \Gamma \text{diag} \left(\|\Lambda(k_1, k_2)\|^2 \right)_{(k_1, k_2) \in L_{[0]}} \Gamma' x_{[0]} \end{aligned} \quad (4.20)$$

where $\Gamma = \left[e_{(k_1, k_2)}(i_1, i_2) \right]_{(i_1, i_2) \in L_{[0]}, (k_1, k_2) \in L_{[0]}}$, which is P with the first row and column removed. The last equality in (4.20) follows from $\|\Lambda(0, 0)\| = 0$ and $x_{(0)}(0, 0) = 0$. Hence

$A'A = \Gamma \text{diag}\left(\|\Lambda(k_1, k_2)\|^2\right)_{(k_1, k_2) \in L \setminus (0,0)} \Gamma'$. It is easy to verify that $\det(\Gamma\Gamma) = 1$, so $\det(A'A) = \prod_{(k_1, k_2) \in L_{\{0\}}} \|\Lambda(k_1, k_2)\|^2$.

Covariance Structure. The representation given in Theorem 4.1 provides a method for computing the covariance structure of the random field. For $(i_1, i_2), (j_1, j_2) \in L$, since $U(k_1, k_2)w_1^{j_1 k_1} w_2^{j_2 k_2} \stackrel{d}{=} U(k_1, k_2)$, using (4.12),

$$Z(i_1, i_2) - Z(j_1, j_2) \stackrel{d}{=} \frac{1}{\sqrt{n_1 n_2}} \sum_{(k_1, k_2) \in L_{\{0\}}} \frac{U(k_1, k_2)}{\|\Lambda(k_1, k_2)\|} \left(w_1^{(i_1 - j_1)k_1} w_2^{(i_2 - j_2)k_2} - 1 \right)$$

Hence,

$$\text{var}(Z(i_1, i_2) - Z(j_1, j_2)) = r(i_1 - j_1, i_2 - j_2) \quad (4.21)$$

where

$$r(i_1, i_2) = \frac{8}{n_1 n_2} \sum_{(k_1, k_2) \in L_{\{0\}}} \frac{1}{\|\Lambda(k_1, k_2)\|^2} \sin^2\left(\pi\left(\frac{i_1 k_1}{n_1} + \frac{i_2 k_2}{n_2}\right)\right) \quad (4.22)$$

Note. The increments are stationary, i.e., the difference variance is a function only of differences $(i_1 - j_1, i_2 - j_2)$. In geostatistics the function r is known as the *variogram*.

Let $X = \text{Re}(Z)$. Covariances for X are obtained using $X(0,0) = 0$,

$$\text{var}(X(i_1, i_2) - X(j_1, j_2)) = \frac{1}{2} \text{var}(Z(i_1, i_2) - Z(j_1, j_2))$$

and

$$\text{cov}(X(i_1, i_2), X(j_1, j_2)) = \frac{1}{2} \left(\text{var}(X(i_1, i_2)) + \text{var}(X(j_1, j_2)) - \text{var}(X(i_1, i_2) - X(j_1, j_2)) \right)$$

which results in

$$\text{cov}(X(i_1, i_2), X(j_1, j_2)) = \frac{1}{4} \left(r(i_1, i_2) + r(j_1, j_2) - r(i_1 - j_1, i_2 - j_2) \right) \quad (4.23)$$

Hence, the variogram completely characterizes the process.

Standard Regularization Model

Hereafter a specific form is assumed for the coefficients a_{kl} of T . With σ and ρ strictly positive, the special form of T is chosen so that

$$\|\Lambda(k_1, k_2)\|^2 = \frac{1}{\sigma^2} \left(|1 - w_1^{-k_1}|^2 + |\rho(1 - w_2^{-k_2})|^2 \right)^p = \frac{4^p}{\sigma^2} \left(\sin^2\left(\frac{\pi k_1}{n_1}\right) + \rho^2 \sin^2\left(\frac{\pi k_2}{n_2}\right) \right)^p \quad (4.24)$$

which is equivalent to

$$\begin{aligned} T_{(1,\sigma,\rho)}x &= \frac{1}{\sigma} (t_1x, \rho t_2x) \\ T_{(2,\sigma,\rho)}x &= \frac{1}{\sigma} (t_1^2x, \sqrt{2}\rho t_1t_2x, \rho^2 t_2^2x) \\ T_{(3,\sigma,\rho)}x &= \frac{1}{\sigma} (t_1^3x, \sqrt{3}\rho t_1^2t_2x, \sqrt{3}\rho^2 t_1t_2^2x, \rho^3 t_2^3x) \end{aligned} \quad (4.25)$$

and so on; coefficients are given by Pascal's triangle, and subscripts denote order and parameter value. With T so defined, the resulting distribution is said here to be of *standard form*.

Note. The pair $\omega = \left(\frac{\pi k_1}{n_1}, \frac{\pi k_2}{n_2}\right)$ is a spatial frequency. With $\sin x \approx x$,

$\|\omega\|^2 \approx \sin^2 \omega_1 + \sin^2 \omega_2$. Hence, the Fourier coefficients $U/\|\Lambda\|$ of the representation (4.12) for small $\|\omega\|$ fall off as roughly $1/\|\omega\|^{p/2}$. This is the law for the "1/f" noises of Mandelbrot [6]. The case $\rho \neq 1$ allows for ellipsoidal falloff. With the above form (4.25) for T the corresponding prior energy or regularization term $H(x) = \|Tx\|^2$ is commonly used in image processing applications [7]. For example, the first order energy enforces smoothness of x by penalizing large gradients. The smoothness increases with order, as high spatial frequencies are increasingly damped in the Fourier representation (4.12).

Samples from the $\rho = 1$ distributions, generated using (4.12), are shown in Figs. 4.1, 4.2, 4.3, 4.4, for orders $p = 1, 2, 3, 4$, respectively. Mesh plots for orders $p = 1, 2, 3$ are

shown in Figs. 4.5, 4.6, 4.7, respectively. As seen in Fig. 4.6, the sample from the second order distribution has the visual characteristics of the height function for a planetary surface. For this reason the second order distribution is used as the surface model in the Bayesian formulation of radarclinometry presented in Chapter 3.

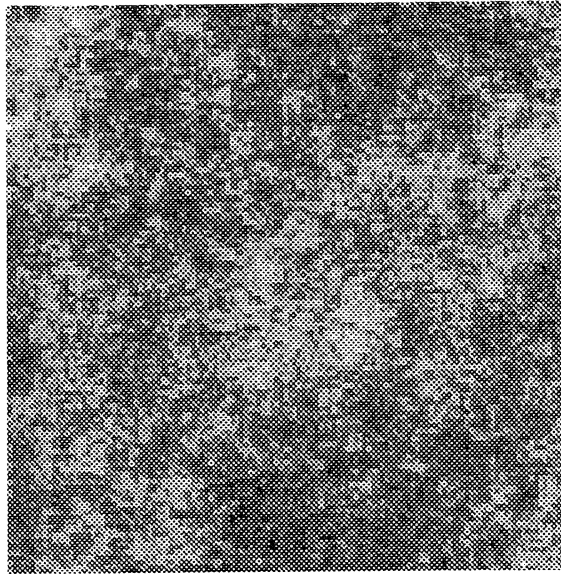


Fig. 4.1. Sample from first order distribution.

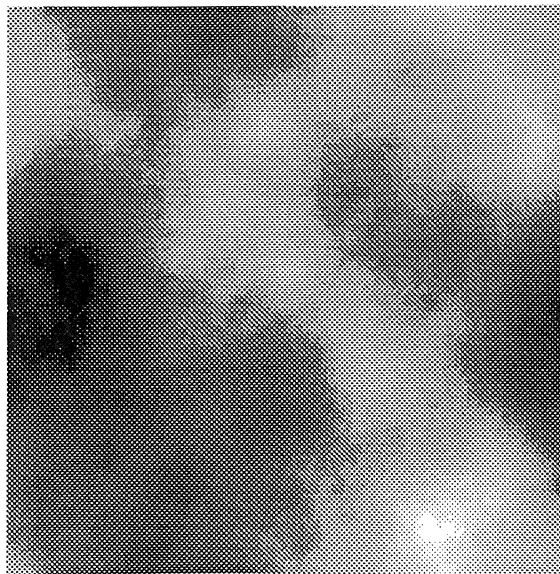


Fig. 4.2. Sample from second order distribution.



Fig. 4.3. Sample from third order distribution.

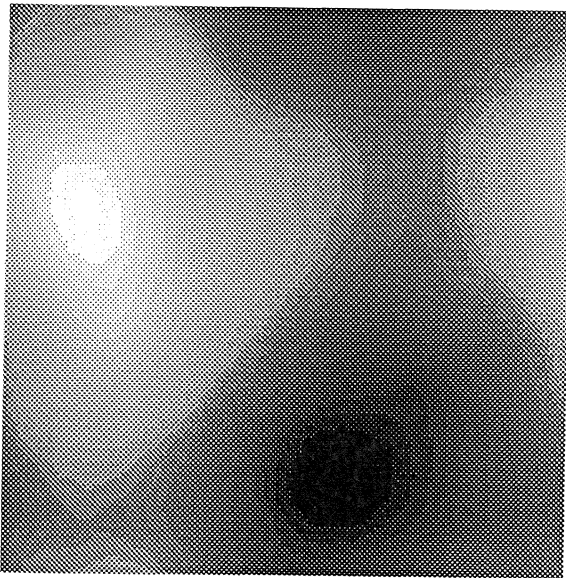


Fig. 4.4. Sample from fourth order distribution.

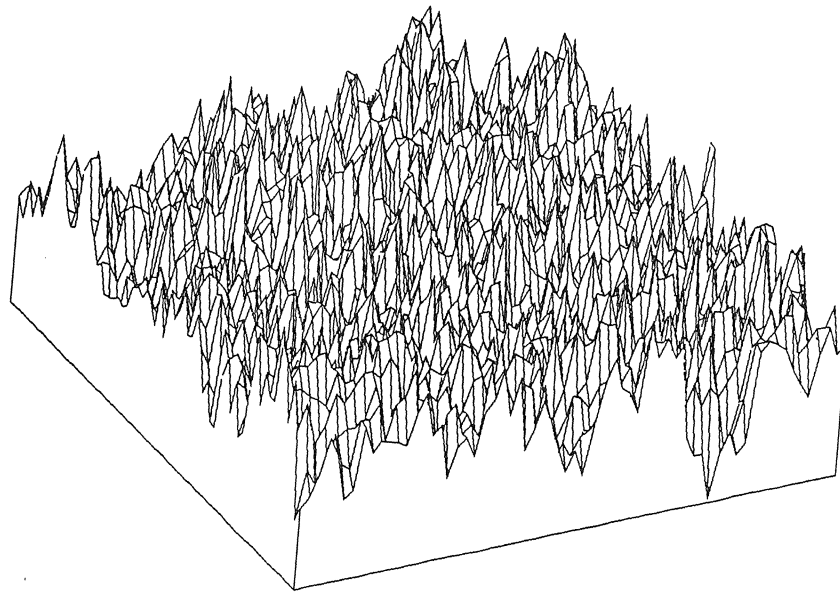


Fig. 4.5. Mesh plot of first order sample.

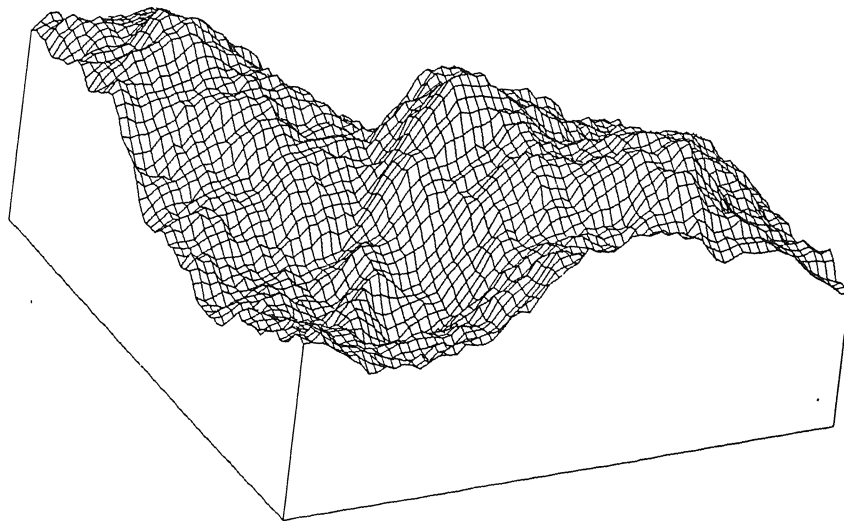


Fig. 4.6. Mesh plot of second order sample.

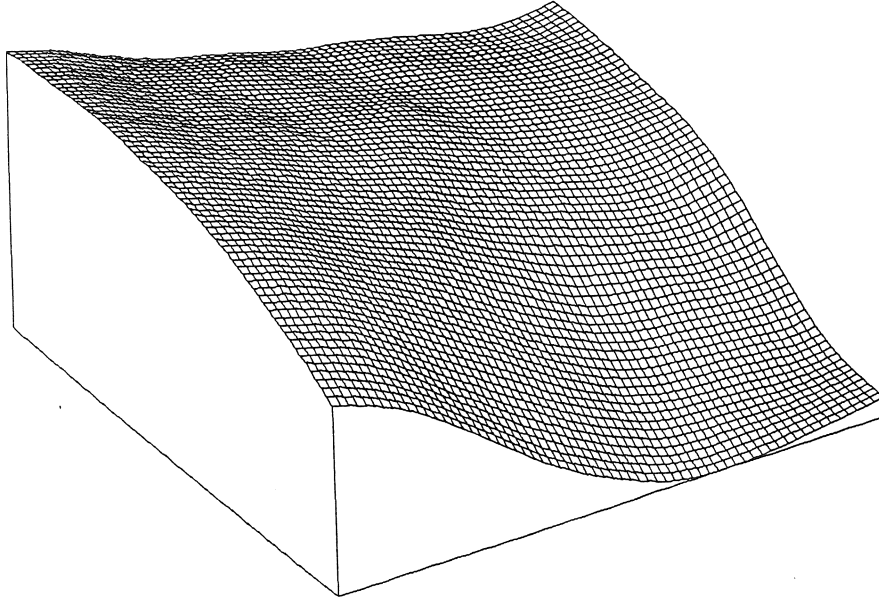


Fig. 4.7. Mesh plot of third order sample.

Alternate Representation. For stabilizing distributions in standard form and of even order, the representation of Theorem 4.1 can be based on the discrete Laplacian. Using (4.19),

$$\begin{aligned}
 \|Tx\|^2 &= \sum_{(k_1, k_2) \in L} \|\Lambda(k_1, k_2)\|^2 |\hat{x}(k_1, k_2)|^2 \\
 &= \sum_{(k_1, k_2) \in L} \frac{1}{\sigma^2} \left(|1 - w_1^{-k_1}|^2 + |\rho(1 - w_2^{-k_2})|^2 \right)^p |\hat{x}(k_1, k_2)|^2 \\
 &= \frac{1}{\sigma^2} x' PD^p P' x = \frac{1}{\sigma^2} x' PD^{p/2} P' PD^{p/2} P' x = \frac{1}{\sigma^2} \|PD^{p/2} P' x\|^2 = \frac{1}{\sigma^2} \|\Delta^{p/2} x\|^2 \quad (4.26)
 \end{aligned}$$

where $D = \text{diag} \left(|1 - w_1^{-k_1}|^2 + |\rho(1 - w_2^{-k_2})|^2 \right)_{(k_1, k_2) \in L}$, and

$$\begin{aligned}
 \Delta x(i_1, i_2) &= x(i_1 + 1, i_2) - 2x(i_1, i_2) + x(i_1 - 1, i_2) \\
 &\quad + \rho^2 (x(i_1, i_2 + 1) - 2x(i_1, i_2) + x(i_1, i_2 - 1)) \quad (4.27)
 \end{aligned}$$

which is the discrete Laplacian for $\rho = 1$. The last equality in (4.26) is verified by noting that the diagonal elements of D are the eigenvalues of $-\Delta$ relative to the Fourier basis. In

other words, $(\text{dft}(-\Delta x))(k_1, k_2) = \lambda(k_1, k_2)\hat{x}(k_1, k_2)$, where
 $\lambda(k_1, k_2) = \left|1 - w_1^{-k_1}\right|^2 + \left|\rho(1 - w_2^{-k_2})\right|^2$. Hence, for the standard form and even orders
 (when $\Delta^{p/2}$ can be defined through composition) the projection (4.13) can be based on
 the Laplacian, i.e.,

$$X \stackrel{d}{=} \text{Re arg min}_{z(0):L \rightarrow \mathbb{C}} \left\| \frac{1}{\sigma} \Delta^{p/2} z(0) - W \right\|^2 \quad (4.28)$$

where W for any order has outcomes in $\mathbb{C}^{n_1 n_2}$. The equivalence (4.28) is true by (4.26)
 and Lemma 4.3.

The representation of Theorem 4.1 is based on knowing the eigenvalues and
 eigenvectors of the operator T . Using (4.28), the analogue of Lemma 4.2 for the operator
 $\Delta^{p/2}$ results in the same representation (4.12), which was based on the operator T . This is
 true because the eigenvectors are Fourier for both operators, and
 $\frac{1}{\sigma} \lambda(k_1, k_2)^{p/2} = \|\Lambda(k_1, k_2)\|$ (see (4.24)). In the next section a modification of the
 Laplacian for which eigenvectors and eigenvalues are known is used to relax the toroidal
 boundary conditions brought about by periodic extension.

Boundary Conditions. Toroidal boundary conditions are sometimes restrictive; for
 example, in image processing applications, where dependence should not “wrap around”
 at boundaries. To obtain alternative boundary conditions for stabilizing distributions in
 standard form and of even order, modify the Laplacian in the following way. On
 $\{(i_1, i_2): i_1 \neq 0, n_1 - 1, i_2 \neq 0, n_2 - 1\}$ let $\bar{\Delta} = \Delta$, but for $i_1 = 0, n_1 - 1$ let

$$\begin{aligned} \bar{\Delta}x(0, i_2) &= \Delta x(0, i_2) - x(n_1 - 1, i_2) + x(0, i_2) \\ \bar{\Delta}x(n_1 - 1, i_2) &= \Delta x(n_1 - 1, i_2) - x(0, i_2) + x(n_1 - 1, i_2) \end{aligned} \quad (4.29)$$

respectively, and similarly for the other coordinate, and carry out both modifications for corner sites. This removes the bonds that wrap around the boundary. As pointed out by Künsch [5], for $(k_1, k_2) \in L_{\{0\}}$, $\bar{\Delta}$ has the eigenvalue

$$\bar{\lambda}(k_1, k_2) = 4(\sin^2(\frac{\pi k_1}{2n_1}) + \rho^2 \sin^2(\frac{\pi k_2}{2n_2}))$$

relative to the eigenvector

$$\bar{e}_{(k_1, k_2)} = \left(2(n_1 n_2)^{-1/2} \cos(\frac{\pi k_1}{n_1}(i_1 + \frac{1}{2})) \cos(\frac{\pi k_2}{n_2}(i_2 + \frac{1}{2})) \right)_{(i_1, i_2) \in L}$$

and, as above, an eigenvalue of zero for the eigenvector $\bar{e}_{(0,0)} = \left((n_1 n_2)^{-1/2} \right)_{(i_1, i_2) \in L}$.

Hence, a complex random field \bar{Z} with the modified standard stabilizing Gibbs distribution has representation

$$\bar{Z}(i_1, i_2) = \sigma \sum_{(k_1, k_2) \in L_{\{0\}}} \frac{U(k_1, k_2)}{\bar{\lambda}(k_1, k_2)^{p/2}} (\bar{e}_{(k_1, k_2)}(i_1, i_2) - \bar{e}_{(k_1, k_2)}(0, 0)) \quad (4.30)$$

where U is distributed as above.

To show the effect of the boundary conditions, samples from the second order distribution, with both periodic and free boundary conditions, are shown in Fig. 4.8. The sample for the periodic conditions is an outcome of $\text{Re}(Z)$ where Z has the distribution of V in (4.12). For the free boundary the sample is an outcome of $\text{Re}(\bar{Z})$ as given in (4.30). As might be expected, local characteristics are similar but the free boundary sample is more “ragged” at the boundaries. Fig. 4.9 shows the zoomed central portions of the images of Fig. 4.8. The figure shows visual similarity (in distribution) of the samples if attention is focused on interiors, away from boundaries. Figures 4.10 and 4.11 show mesh plots of the images of Fig. 4.9.

In a different context the use of a prior based on the Laplacian and its eigenvalue/eigenvector representation can be found in Amit *et al.* [1].

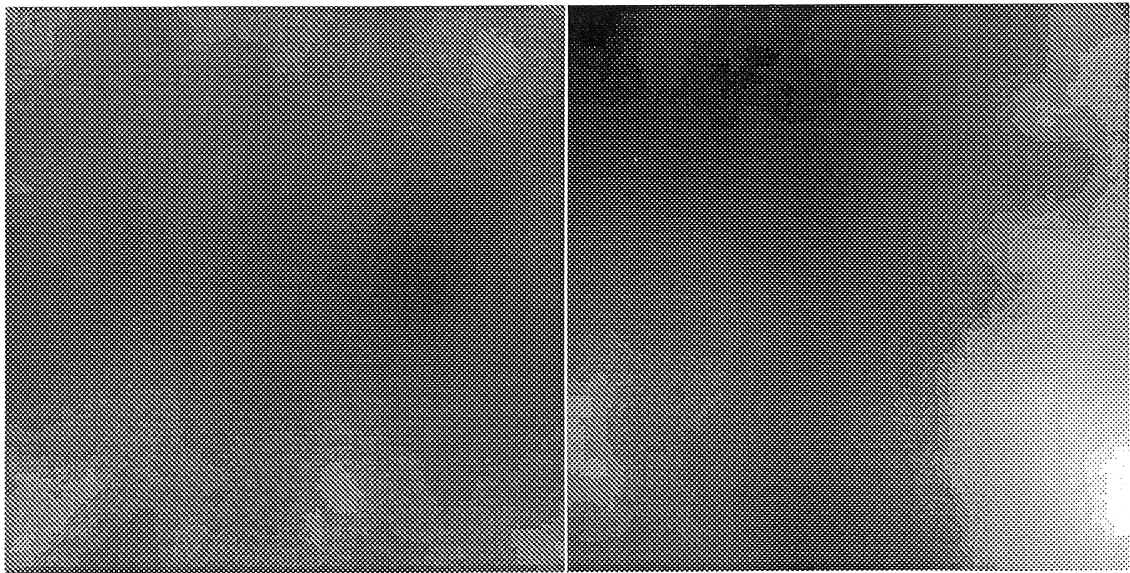


Fig. 4.8. Samples from second order distribution;
left: periodic boundary; right: free boundary.

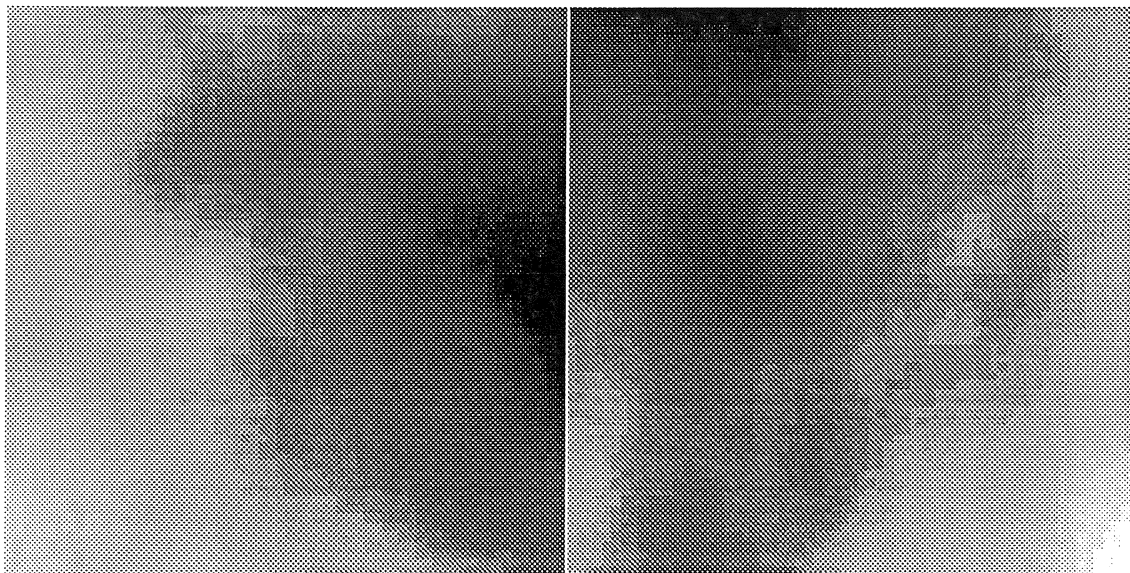


Fig. 4.9. Central portions of Fig. 4.5.

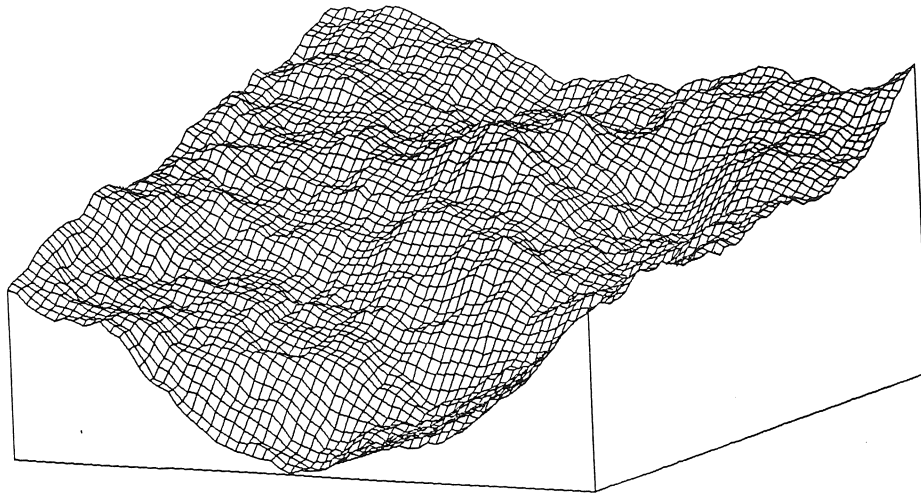


Fig. 4.10. Mesh plot, periodic boundary.

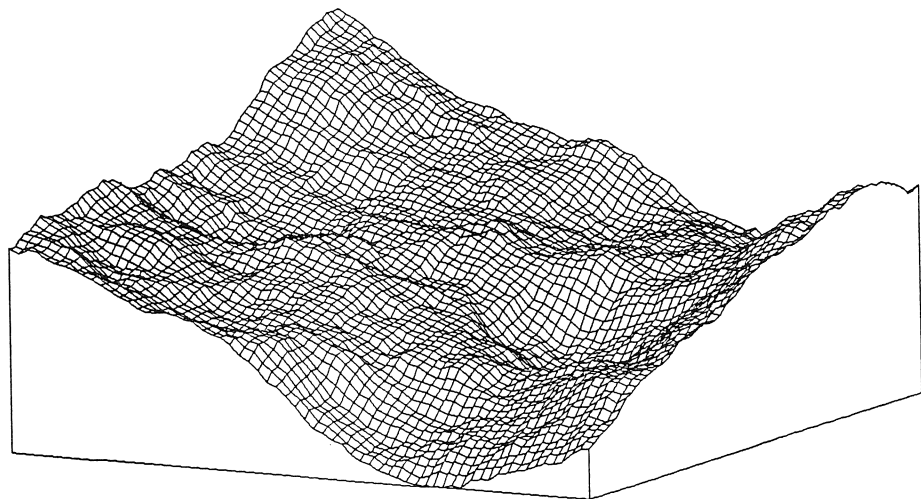


Fig. 4.11. Mesh plot, free boundary.

ML Estimate. Assume the scaling parameter σ is unknown in the standard model. Let x be a member of the equivalence class for an outcome of X . Using the normalized form of the stabilizing Gibbs distribution, the likelihood function for σ^2 can be written

$$l(\sigma^2, x) = -(n_1 n_2 - 1) \log(\sigma^2) - \frac{1}{\sigma^2} \|\sigma T x\|^2 \quad (4.31)$$

so the maximum likelihood estimate for σ^2 is simply $\frac{\|\sigma T x\|^2}{n_1 n_2 - 1}$.¹ For p even, the identity (4.26) leads to a simplified computation of the quadratic.

Self-similarity Property

Self-similarity is the property that a scaling of coordinates is equivalent to a scaling of the process (Mandelbrot [5]). In other words, if a sample from a self-similar process is “zoomed in” the result looks similar to the original process. Self-similar processes have “similar characteristics over a range of scales.” This is considered an important property of the $1/f$ noises of Mandelbrot because self-similarity is claimed to be characteristic of many natural phenomena.

It is not surprising that the stabilizing Gibbs distribution of standard form obeys a self-similarity property given the aforementioned relationship to $1/f$ noises. However, only an asymptotic statement can be made, and it is valid only for orders $p \geq 2$. In any event, the stabilizing Gibbs distribution is used here to model natural surfaces, which is appropriate given the claim that self-similarity occurs in nature. Synthetic scenes generated from $1/f$ noises are shown in [6] to support this claim.

The important distinction between the stabilizing Gibbs distribution and $1/f$ noises for modeling surfaces is that the random field is directly defined on a lattice (as opposed to the plane) and has a local neighborhood structure. These are important computational considerations in the context of stochastic inverse problems. Further, as is

¹By (4.25), the numerator does not depend on σ .

described in detail later, there is another computational benefit due to self-similarity: the property suggests a procedure for maintaining consistency among levels in a multi-resolution hierarchy.

The basic result is presented next. In the following, let $L^{(n_1, n_2)} = \{0, 1, \dots, n_1 - 1\} \times \{0, 1, \dots, n_2 - 1\}$ and $L_{[0]}^{(n_1, n_2)} = L^{(n_1, n_2)} \setminus (0, 0)$.

Corollary 4.1. Let a_1, a_2 be integers greater than or equal to one, and let the order $p \geq 2$. Let $\sigma' = a_1^{-p+1/2} a_2^{1/2} \sigma$ and $\rho' = (a_2/a_1)\rho$. Fix (n'_1, n'_2) . Then the variogram obeys

$$\frac{1}{\Phi_{n_1, n_2}} \left| r_{(\sigma, \rho)}^{(n_1, n_2)}(l_1, l_2) - r_{(\sigma', \rho')}^{(a_1 n_1, a_2 n_2)}(a_1 l_1, a_2 l_2) \right| \xrightarrow{n_1, n_2 \rightarrow \infty} 0$$

uniformly for all $(l_1, l_2) \neq 0$, $|l_1| < n'_1$, $|l_2| < n'_2$, where for r superscripts denote lattice size, and $\Phi_{n_1, n_2} = r_{(1, \rho)}^{(n_1, n_2)}(n'_1/2, n'_2/2)$, a normalizing constant which ensures that $\Phi_{n_1, n_2}^{-1} r_{(\sigma, \rho)}^{(n_1, n_2)}(n'_1/2, n'_2/2) = \sigma^2$ for all (n_1, n_2) .

Application. Let X and X' have stabilizing Gibbs distributions of order p , with sizes (n_1, n_2) and $(a_1 n_1, a_2 n_2)$, parameters (σ, ρ) and $(a_1^{-p+1/2} a_2^{1/2} \sigma, (a_2/a_1)\rho)$, respectively. The corollary together with (4.23) implies that

$$\Phi_{n_1, n_2}^{-1/2} (X(i_1, i_2))_{(i_1, i_2) \in L^{(n_1, n_2)}} \approx \Phi_{n_1, n_2}^{-1/2} (X'(a_1 i_1, a_2 i_2))_{(i_1, i_2) \in L^{(n_1, n_2)}} \quad (4.32)$$

in distribution. The processes converge in distribution to each other on a fixed sublattice $L^{(n'_1, n'_2)}$. However, the corollary states that for small spatial lags (l_1, l_2) the variograms converge to one another, which is a statement about small-scale properties for all of $L^{(n_1, n_2)}$. Thus, in a sense (admittedly weak), (4.32) holds for the processes defined on $L^{(n_1, n_2)}$.

Define $G_{(\Delta_1, \Delta_2)}^{(n_1, n_2)} = \{(i_1 \Delta_1, i_2 \Delta_2) : (i_1, i_2) \in L^{(n_1, n_2)}\}$, a grid with spacing (Δ_1, Δ_2) . Associate X with the grid $G_{(\Delta_1, \Delta_2)}^{(n_1, n_2)}$ and X' with the grid $G_{(\Delta_1/a_1, \Delta_2/a_2)}^{(a_1 n_1, a_2 n_2)}$. Thus

$(X(i_1, i_2))_{(i_1, i_2) \in L^{(n_1, n_2)}}$ and $(X'(a_1 i_1, a_2 i_2))_{(i_1, i_2) \in L^{(n_1, n_2)}}$ “correspond” to one another with respect to the grids. The marginal distribution of X' on the grid sites on which X is defined is approximately equal to the distribution of X . This equivalence is important in maintaining consistency of the stabilizing Gibbs distribution in the multi-resolution scheme described later in Chapter 6.

Example. Let $p = 2$, $a_1 = a_2 = 2$. Then X has approximately the same distribution as $\frac{1}{2}X'$ on its even numbered sites.

Self-similarity. Let $a = a_1 = a_2$, and let X and X' have stabilizing Gibbs distributions of order p , common parameter (σ, ρ) , and sizes (n_1, n_2) and $(a_1 n_1, a_2 n_2)$, respectively. One then has the self-similarity property

$$a^{p-1} (X(i_1, i_2))_{(i_1, i_2) \in L^{(n_1, n_2)}} \approx (X'(a i_1, a i_2))_{(i_1, i_2) \in L^{(n_1, n_2)}} \quad (4.33)$$

in distribution, i.e., a scaling of the spatial coordinates is equivalent to scaling the process. Fig. 4.12 shows the variograms $r_{(\sigma, 1)}^{(128, 128)}$ and $r_{(2\sigma, 1)}^{(64, 64)}$ for comparison, which is strongly suggestive of the approximate equivalence in distribution (on $L^{(n_1, n_2)}$). The origin is at the center in both cases.

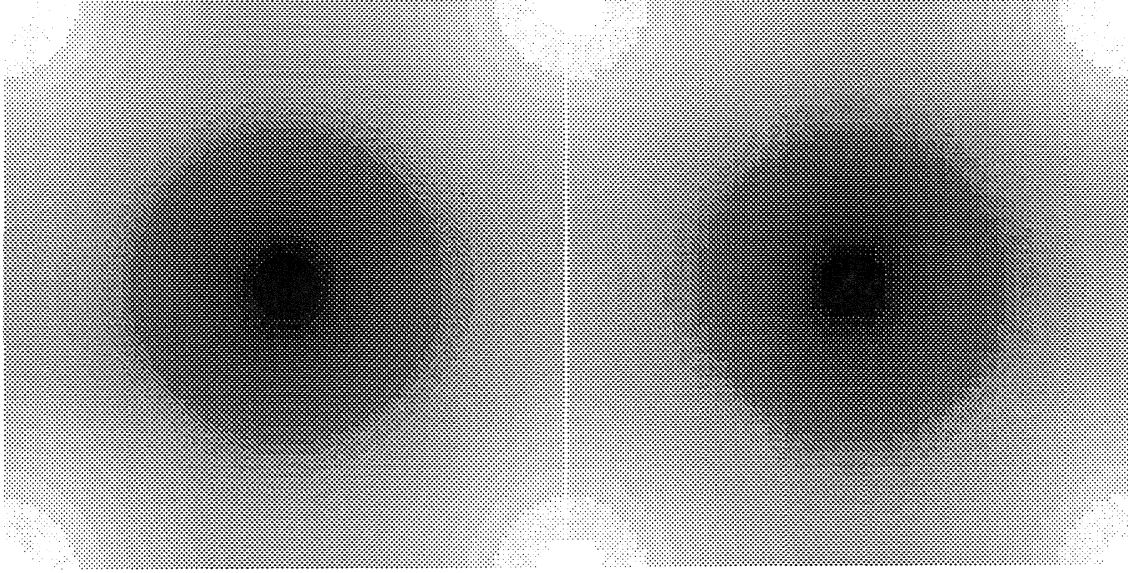


Fig. 4.12. Difference variances.
left: 128×128 ; right: 64×64 .

Proof. The method of proof is to bound $\left| r_{(\sigma, \rho)}^{(n_1, n_2)}(l_1, l_2) - r_{(\sigma', \rho')}^{(a_1 n_1, a_2 n_2)}(a_1 l_1, a_2 l_2) \right|$ from above and φ_{n_1, n_2} from below. Proceeding with the former, using (4.22),

$$\begin{aligned}
 & \left| r_{(\sigma, \rho)}^{(n_1, n_2)}(l_1, l_2) - r_{(\sigma', \rho')}^{(a_1 n_1, a_2 n_2)}(a_1 l_1, a_2 l_2) \right| \\
 &= \left| \frac{2\sigma^2}{4^{p-1} n_1 n_2} \sum_{(k_1, k_2) \in L_{[0]}^{(n_1, n_2)}} \frac{\sin^2(\pi(l_1 k_1/n_1 + l_2 k_2/n_2))}{\left(\sin^2(\pi k_1/n_1) + \rho^2 \sin^2(\pi k_2/n_2)\right)^p} \right. \\
 & \quad \left. \frac{2a_1^{1-2p} a_2 \sigma^2}{4^{p-1} a_1 a_2 n_1 n_2} \sum_{(k_1, k_2) \in L_{[0]}^{(a_1 n_1, a_2 n_2)}} \frac{\sin^2(\pi(l_1 k_1/n_1 + l_2 k_2/n_2))}{\left(\sin^2(\pi k_1/a_1 n_1) + (a_2/a_1)^2 \rho^2 \sin^2(\pi k_2/a_2 n_2)\right)^p} \right| \\
 &= \left| \frac{2\sigma^2}{4^{p-1}} \frac{1}{n_1 n_2} \sum_{(k_1, k_2) \in L_{[0]}^{(n_1, n_2)}} g(k_1, k_2) \right| \tag{4.34}
 \end{aligned}$$

where

$$g(k_1, k_2) = \sin^2(\pi(l_1 k_1/n_1 + l_2 k_2/n_2))$$

$$\left(\frac{1}{\left(\sin^2(\pi k_1/n_1) + \rho^2 \sin^2(\pi k_2/n_2)\right)^p} - \sum_{(i_1, i_2) \in L^{(a_1, a_2)}} f_{(i_1, i_2)}(k_1, k_2) \right) \quad (4.35)$$

$$f_{(i_1, i_2)}(k_1, k_2) = \frac{1}{\left(a_1^2 \sin^2(\pi(k_1 + i_1 n_1)/a_1 n_1) + a_2^2 \rho^2 \sin^2(\pi(k_2 + i_2 n_2)/a_2 n_2)\right)^p} \quad (4.36)$$

Further,

$$\left| \frac{1}{n_1 n_2} \sum_{(k_1, k_2) \in L_{[0]}^{(n_1, n_2)}} g(k_1, k_2) \right| \leq \quad (4.37)$$

$$\frac{1}{n_1 n_2} \left(\left| \sum_{(k_1, k_2) \in Q_1} g(k_1, k_2) \right| + \left| \sum_{(k_1, k_2) \in Q_2} g(k_1, k_2) \right| + \left| \sum_{(k_1, k_2) \in Q_3} g(k_1, k_2) \right| + \left| \sum_{(k_1, k_2) \in Q_4} g(k_1, k_2) \right| \right)$$

where $Q_1 = [0, n_1/2) \times [0, n_2/2) \setminus (0, 0)$, $Q_2 = [n_1/2, n_1) \times [0, n_2/2)$,

$Q_3 = [0, n_1/2) \times [n_2/2, n_2)$, and $Q_4 = [n_1/2, n_1) \times [n_2/2, n_2)$.

The sum is easy to bound except in the four corners of $[0, n_1) \times [0, n_2)$. Let $\omega_1 = \pi k_1/n_1$, $\omega_2 = \pi k_2/n_2$. To bound the sum in the first quadrant,

$$\left| \frac{1}{n_1 n_2} \sum_{(k_1, k_2) \in Q_1} g(k_1, k_2) \right|$$

$$\leq \frac{1}{n_1 n_2} \sum_{(k_1, k_2) \in Q_1} \left| \sin^2(l_1 \omega_1 + l_2 \omega_2) \left(\frac{1}{\left(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2\right)^p} - f_{(0,0)}(k_1, k_2) \right) \right| \quad (4.38)$$

$$+ \frac{1}{n_1 n_2} \sum_{(k_1, k_2) \in Q_1} \sum_{(i_1, i_2) \in L_{[0]}^{(a_1, a_2)}} |f_{(i_1, i_2)}(k_1, k_2)|$$

First bound the second sum on the right-hand side as follows. For $a, n \geq 1$, $k \in [0, n/2)$, and $l \in [1, a)$,

$$1/a \leq (k + ln)/an = k/an + l/a \leq (n/2)/an + (a-1)/a = 1 - 1/2a \quad (4.39)$$

so $\sin^2(\pi(k + ln)/an) \geq \sin^2(\pi/2a)$. Applying this to the above, for $(k_1, k_2) \in Q_1$ and $(i_1, i_2) \in L_{[0]}^{(a_1, a_2)}$,

$$f_{(i_1, i_2)}(k_1, k_2) \leq \frac{1}{\left(\min(a_1^2, a_2^2 \rho^2) \sin^2(\pi/2 \max(a_1, a_2))\right)^p} \quad (4.40)$$

so the sum is bounded by a constant. It remains to bound the first sum on the right-hand side above. First bound the summand as follows.

$$\begin{aligned} & \left| \sin^2(l_1 \omega_1 + l_2 \omega_2) \left(\frac{1}{\left(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2\right)^p} - f_{(0,0)}(k_1, k_2) \right) \right| \\ &= \sin^2(l_1 \omega_1 + l_2 \omega_2) \left| \left(\frac{1}{\left(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2\right)^p} - \frac{1}{\left(a_1^2 \sin^2(\omega_1/a_1) + a_2^2 \rho^2 \sin^2(\omega_2/a_2)\right)^p} \right) \right| \end{aligned} \quad (4.41)$$

Using the expansion $\sin^2 x = x^2 + e(x)$, where $\lim_{x \rightarrow 0} \frac{e(x)}{x^4} = -12$,

$$\begin{aligned} & \frac{1}{\left(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2\right)^p} - \frac{1}{\left(a_1^2 \sin^2(\omega_1/a_1) + a_2^2 \rho^2 \sin^2(\omega_2/a_2)\right)^p} \\ &= \frac{\left(a_1^2 \sin^2(\omega_1/a_1) + a_2^2 \rho^2 \sin^2(\omega_2/a_2)\right)^p - \left(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2\right)^p}{\left(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2\right)^p \left(a_1^2 \sin^2(\omega_1/a_1) + a_2^2 \rho^2 \sin^2(\omega_2/a_2)\right)^p} \\ &= \frac{\left(\omega_1^2 + \rho^2 \omega_2^2 + a_1^2 e(\omega_1/a_1) + a_2^2 \rho^2 e(\omega_2/a_2)\right)^p - \left(\omega_1^2 + \rho^2 \omega_2^2 + e(\omega_1) + \rho^2 e(\omega_2)\right)^p}{\left(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2\right)^p \left(a_1^2 \sin^2(\omega_1/a_1) + a_2^2 \rho^2 \sin^2(\omega_2/a_2)\right)^p} \end{aligned}$$

and by the binomial theorem, using $e(x) = O(x^4)$,

$$\begin{aligned}
&= \frac{\sum_{i=1}^p \binom{p}{i} (\omega_1^2 + \rho^2 \omega_2^2)^{p-i} \left((a_1^2 e(\omega_1/a_1) + a_2^2 \rho^2 e(\omega_2/a_2))^i - (e(\omega_1) + \rho^2 e(\omega_2))^i \right)}{(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2)^p (a_1^2 \sin^2(\omega_1/a_1) + a_2^2 \rho^2 \sin^2(\omega_2/a_2))^p} \quad (4.42) \\
&= \frac{O(\|\omega\|^{2(p-1)})O(\|\omega\|^4)}{O(\|\omega\|^{2p})O(\|\omega\|^{2p})} = O(\|\omega\|^{2-2p})
\end{aligned}$$

where $\omega = (\omega_1, \omega_2)$, and $f(\omega_1, \omega_2) = O(\|\omega\|^n)$ denotes the fact that there exist constants $m, M > 0$ such that $m \leq \frac{|f(\omega_1, \omega_2)|}{\|\omega\|^n} \leq M$ for all $\|\omega\|$ sufficiently small. Now, with

$\sin^2(l_1\omega_1 + l_2\omega_2) = O(\|\omega\|^2)$, it is evident that there exists an $\varepsilon > 0$ such that for $\|\omega\| < \varepsilon$,

$$\left| \sin^2(l_1\omega_1 + l_2\omega_2) \left(\frac{1}{(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2)^p} - f_{(0,0)}(k_1, k_2) \right) \right| \leq \text{const.} \|\omega\|^{4-2p} \quad (4.43)$$

Hence,

$$\begin{aligned}
&\frac{1}{n_1 n_2} \sum_{(k_1, k_2) \in Q_1} \left| \sin^2(l_1\omega_1 + l_2\omega_2) \left(\frac{1}{(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2)^p} - f_{(0,0)}(k_1, k_2) \right) \right| \\
&\leq \text{const.} \frac{1}{n_1 n_2} \sum_{\{(k_1, k_2) \in Q_1 : \|\omega\| < \varepsilon\}} \|\omega\|^{4-2p} \\
&+ \frac{1}{n_1 n_2} \sum_{\{(k_1, k_2) \in Q_1 : \|\omega\| \geq \varepsilon\}} \left| \left(\frac{1}{(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2)^p} - \frac{1}{(a_1^2 \sin^2(\omega_1/a_1) + a_2^2 \rho^2 \sin^2(\omega_2/a_2))^p} \right) \right| \quad (4.44)
\end{aligned}$$

The second sum in (4.44) is clearly bounded by a constant, as is the first for $p = 2$. The first term of (4.44) is a Riemann sum, and for all n_1, n_2 and $p \geq 3$, by definition of the Riemann integral,

$$\begin{aligned}
\frac{1}{n_1 n_2} \sum_{\{(k_1, k_2) \in Q_1: \|\omega\| < \varepsilon\}} \|\omega\|^{4-2p} &\leq \int_{\tan^{-1} \frac{\mu_2}{\mu_1} \in [0, \pi/2), \zeta(n_1, n_2) < \|\mu\| < \varepsilon} \|\mu\|^{4-2p} du_1 du_2 = \pi/2 \int_{\zeta(n_1, n_2) < r < \varepsilon} r^{5-2p} dr \\
&= \begin{cases} -\frac{\pi}{2} (\log \zeta(n_1, n_2) - \log \varepsilon) & p = 3 \\ \frac{\pi}{2(2p-6)} (\zeta^{6-2p}(n_1, n_2) - \varepsilon^{6-2p}) & p = 4, 5, \dots \end{cases} \quad (4.45)
\end{aligned}$$

where $\zeta(n_1, n_2) = \frac{\pi}{2 \max(n_1, n_2)}$; clearly $\zeta(n_1, n_2) \xrightarrow{n_1, n_2 \rightarrow \infty} 0$.

The other three quadrants can be bounded similarly. Note that in all cases an expansion about the origin may be used due to periodicity; there are copies of the remaining three corners at the origin. The following summarizes the upper bounds established thus far.

$$\begin{aligned}
&\left| r_{(\sigma, \rho)}^{(n_1, n_2)}(l_1, l_2) - r_{(\sigma', \rho')}^{(a_1 n_1, a_2 n_2)}(a_1 l_1, a_2 l_2) \right| \\
&\leq \begin{cases} \text{const.} & p = 2 \\ \text{const.} \log(\max(n_1, n_2)) + \text{const.} & p = 3 \\ \text{const.} (\max(n_1, n_2))^{2p-6} + \text{const.} & p = 4, 5, \dots \end{cases} \quad (4.46)
\end{aligned}$$

The normalizing function φ_{n_1, n_2} is bounded from below in the following.

$$\begin{aligned}
\varphi_{n_1, n_2} &= r_{(1, \rho)}^{(n_1, n_2)}(n'_1/2, n'_2/2) \\
&= \frac{2}{4^{p-1} n_1 n_2} \sum_{(k_1, k_2) \in L_{[0]}^{(n_1, n_2)}} \frac{\sin^2(n'_1 \omega_1/2 + n'_2 \omega_2/2)}{(\sin^2 \omega_1 + \rho^2 \sin^2 \omega_2)^p} \\
&= \frac{1}{n_1 n_2} \sum_{(k_1, k_2) \in L_{[0]}^{(n_1, n_2)}} O(\|\omega\|^{2-2p})
\end{aligned} \quad (4.47)$$

and for some $\delta > 0$ and $\gamma(n_1, n_2) = \frac{2\pi}{\min(n_1, n_2)}$,

$$\begin{aligned}
&\geq \frac{\text{const.}}{n_1 n_2} \sum_{(k_1, k_2) \in [0, n_1] \times [0, n_2] \setminus (0, 0), \|\omega\| < \delta} \|\omega\|^{2-2p} \geq \text{const.} \int_{\gamma(n_1, n_2) < r < \delta} r^{3-2p} dr \\
&= \begin{cases} \text{const.} \log(\min(n_1, n_2)) + \text{const.} & p = 2 \\ \text{const.} (\min(n_1, n_2))^{2p-4} + \text{const.} & p = 3, 4, \dots \end{cases} \quad (4.48)
\end{aligned}$$

The result now follows.

References

1. Y. Amit, U. Grenander, and M. Piccioni, (1991). "Structural image restoration through deformable templates," *J. Amer. Stat. Assoc.*, 86, pp. 376-387.
2. J. Besag, (1974). "Spatial interaction and the statistical analysis of lattice systems," *J. Roy. Statist. Soc. B*, 36, pp. 192-236.
3. N. R. Goodman, (1963). "Statistical analysis based on a certain multivariate complex Gaussian distribution," *Ann. Math Statist.*, 34, pp. 152-176.
4. H. R. Künsch, (1987). "Intrinsic autoregressions and related models on the two-dimensional lattice," *Biometrika*, 74, pp. 517-524.
5. H. R. Künsch, (1992). Personal communication.
6. B. B. Mandelbrot, (1983). *The Fractal Geometry of Nature*. New York: W. H. Freeman & Co.
7. T. Poggio, V. Torre, and C. Koch, (1985). "Computational vision and regularization theory," *Nature*, 317, pp. 314-319.
8. B. D. Ripley, (1981). *Spatial Statistics*. New York: John Wiley and Sons.
9. A. N. Tikhonov and V. Y. Arsenin, (1977). *Solutions of Ill-posed Problems*. New York: John Wiley and Sons.

CHAPTER 5

EXTENDED GIBBS SAMPLER

The Gibbs sampler is a Monte Carlo algorithm originally devised for Bayesian methods of image processing (Geman and Geman [5]). In this chapter an extension of this computational technique used in the solution of radarclinometry as a stochastic inverse problem is described. In the extension the Gibbs energy is a sum $H = H_1 + H_2$. A state is proposed by drawing from the standard Gibbs sampler based on H_1 , and accepting based on a Metropolis-type ratio involving H_2 . For radarclinometry, because the prior is normal and the data term is computationally demanding, the method is well suited to the posterior distribution with the regularization term playing the role of H_1 and the data term the role of H_2 .

The results are presented in more generality than is required for the radarclinometry application. The opportunity is taken to extend the Gibbs sampler to include additional attractive image processing and other practical applications. In particular, the Gibbs sampler is extended in two ways: 1) a propose/decide scheme is used, which is a special case of a non-stationary version of the Hastings extension [9] to the Metropolis algorithm [11]; 2) transition operators have general support, in that local updates are based on restrictions of the state space more general than allowing only a change at a single site. A convergence theorem is proved, optimality properties are demonstrated, and some image analysis applications are described.

Algorithms of the type presented here have found wide use in image processing and spatial statistics. For recent expositions see Besag and Green [1] and Tierney [14].

Gibbs Sampler

The standard Gibbs sampler is given as follows. For a finite lattice, L , let $\Omega = \Lambda^L = \{x = (x_s)_{s \in L} : x_s \in \Lambda, s \in L\}$ be a product space based on the lattice, where Λ is a finite set. Define a family, indexed by *temperature* $T > 0$, of probability mass functions on this space, where each element is a Gibbs distribution: for $x \in \Omega$,

$$\Pi_T(x) = \frac{1}{Z_T} \exp(-H(x)/T) \quad (5.1)$$

with H a real-valued *energy function*, and $Z_T = \sum_{x \in \Omega} \exp(-H(x)/T)$ a normalizing

constant. This form is not restrictive; any distribution Π which satisfies $\Pi(x) > 0$ for all $x \in \Omega$ is a Gibbs distribution with $H(x) = -\log \Pi(x)$. Let $(a_i)_{i=0,1,2,\dots}$ be a sequence of sites in the lattice which represents a *site-visitation schedule*, in that i is a time index. Let

$$P_T^{(i)}(x, y) = \Pi_{a_i}(y_{a_i} | x_{(a_i)}) 1_{y_{(a_i)} = x_{(a_i)}} \quad (5.2)$$

where, for $s \in L$,

$$\Pi_s(x_s | x_{(s)}) = \frac{\Pi_T(x)}{\sum_{z: z_{(s)} = x_{(s)}} \Pi_T(z)} \quad (5.3)$$

is the *local distribution at site s* , with $x_{(s)} = (x_t)_{t \neq s, t \in L}$. The results in Geman and Geman

[4] which are relevant here are summarized as follows. 1) Let (X_i) be the non-stationary

Markov chain with transition operators $(P_1^{(i)})$, and for each $s \in L$, suppose $a_i = s$ infinitely often. Then, for any starting point x , $\lim_{i \rightarrow \infty} P(X_i = y | X_0 = x) = \Pi_1(y)$ for all

$y \in \Omega$. 2) Suppose there exists an integer τ such that $L \subseteq \{a_{i+1}, a_{i+2}, \dots, a_{i+\tau}\}$ for all $i \geq 0$.

Fix a *temperature schedule* (T_i) which decreases to zero slowly enough that

$\liminf_{i \rightarrow \infty} T_i \log i$ is sufficiently large. Let (X_i) have transitions $(P_{T_i}^{(i)})$. Then

$\lim_{i \rightarrow \infty} P(X_i = y | X_0 = x) = \Pi_0(y)$, where $\Pi_0(x) = \lim_{T \rightarrow 0} \Pi_T(x)$, which is the distribution

uniform on the states which maximize Π_1 .

The conditions on the site visitation schedule ensure that any state can be reached from any other. The practical significance of the two results is the following. Typically in an image processing setting the local distributions can be easily evaluated, in that through cancellation between numerator and denominator, each calculation involves only a small neighborhood $\eta_s \subset L$ of site s . The first result provides a method of sampling from Π_1 via these local calculations. The second provides a scheme for sampling from Π_0 .

The Gibbs sampler is essentially specified by an energy function and a site visitation schedule. These two components are modified here in such a way that the standard Gibbs specification is a special case.

The energy function determines the distribution from which the Gibbs sampler generates realizations. In the extension given here the energy is a sum of two terms, motivated by the Bayesian setting, as follows. A posterior distribution $P(X = x | Y = y)$ is the object of interest where x is the scene, or desired information sought from the data y . The posterior distribution is Gibbs (assuming positivity) with $H(x) = -\log P(Y = y | X = x) - \log P(X = x)$. Either of these terms can play the role of $H_1(x)$ or $H_2(x)$, but this need not be the case. For example, as in Green and Han [6], H_1 can be an approximation to H , with H_2 the error.

The calculation involved in generating samples from Π_T directly is reduced to sampling from local distributions, and the site visitation schedule gives the order of sampling from these. At each time i , the transition probabilities are restricted, renormalized versions of Π_T , i.e., of the form

$$P_T^{(i)}(x, y) = \frac{\Pi_T(y)}{\sum_{z \in S_i(x)} \Pi_T(z)} 1_{S_i(x)}(y) \quad (5.4)$$

where $S_i(x) = \{z: z_{(a_i)} = x_{(a_i)}\}$. So for each $x \in \Omega$, $S_i(x) \subset \Omega$. This is the motivation for replacing the site visitation schedule (a_i) with an arbitrary sequence (S_i) of subsets of Ω .

Extended Gibbs Sampler

For an arbitrary finite sample space Ω let

$$\Pi_T(x) = \frac{1}{Z_T} \exp(-(H_1(x) + H_2(x))/T) \quad (5.5)$$

be a probability mass function on this space, where $Z_T = \sum_{x \in \Omega} \exp(-(H_1(x) + H_2(x))/T)$.

Define a sequence of maps $(S_i)_{i=0,1,2,\dots}$, where $S_i: \Omega \rightarrow 2^\Omega$. Thus, for $x \in \Omega$ and each i $S_i(x)$ is a subset of Ω . Also, assume the property $y \in S_i(x) \Leftrightarrow x \in S_i(y)$ holds for each $x, y \in \Omega$ and each i . Define the transitions

$$P_T^{(i)}(x, y) = \begin{cases} Q_T^{(i)}(x, y) \min(1, \rho_T^{(i)}(x, y)) & y \neq x \\ 1 - \sum_{z \neq x} Q_T^{(i)}(x, z) \min(1, \rho_T^{(i)}(x, z)) & y = x \end{cases} \quad (5.6)$$

where

$$\begin{aligned} Q_T^{(i)}(x, y) &= \frac{\exp(-H_1(y)/T)}{\sum_{z \in S_i(x)} \exp(-H_1(z)/T)} 1_{S_i(x)}(y) \\ \rho_T^{(i)}(x, y) &= \frac{\exp(-H_2(y)/T) \sum_{z \in S_i(x)} \exp(-H_1(z)/T)}{\exp(-H_2(x)/T) \sum_{z \in S_i(y)} \exp(-H_1(z)/T)} \end{aligned} \quad (5.7)$$

which equals $\frac{\Pi_T(y) Q_T(y, x)}{\Pi_T(x) Q_T(x, y)}$ if $y \in S_i(x)$.

Following Hastings [9], $\rho_T^{(i)}$ is called the *test ratio*. It is easy to verify that these transitions result from the following algorithm, carried out at each time i :

Propose: state y is proposed as a new state with probability $Q_T^{(i)}(x, y)$, where x is the current state

Decide: the new state is accepted with probability $\min(1, \rho_T^{(i)}(x, y))$

The following theorem extends the two results presented earlier for the Gibbs sampler. Basically, the chain has the appropriate limit distribution. The constant

temperature result corresponds to limit distribution $\Pi' = \Pi_1$ below, the simulated annealing result to $\Pi' = \Pi_0$.

Theorem 5.1. Let $(T_k)_{k=0,1,2,\dots}$ be a sequence of positive numbers such that $\Pi_{T_k} \rightarrow \Pi'$, (T_k) is eventually decreasing, and $\liminf_{k \rightarrow \infty} T_k \log k$ is sufficiently large. Suppose there exists a sequence $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \dots$ such that for all $k \geq 0$ and $x, y \in \Omega$, $y \in S_{\tau_{k+1}-1} \circ S_{\tau_{k+1}-2} \circ \dots \circ S_{\tau_k+1} \circ S_{\tau_k}(x)$, and there is an n such that $\tau_{k+1} - \tau_k \leq n \forall k$. This condition ensures that any state can be reached from any other state in finite time. Let $(X_i)_{i=0,1,2,\dots}$ be a Markov chain with transitions $(P_{T_{k_i}}^{(i)})_{i=0,1,2,\dots}$, where $k_i \equiv k$ for $i \in [\tau_k, \tau_{k+1})$. Then

$$\lim_{i \rightarrow \infty} P(X_i = y | X_0 = x) = \Pi'(y) \quad \forall x, y \in \Omega \quad (5.8)$$

Before presenting the proof, some applications are described.

Applications

Existing and new algorithms are shown to be applications of Theorem 5.1. In all cases, including the standard Gibbs sampler, the condition that each state can be reached from any other is satisfied by ensuring that all sites are repeatedly visited. A number of properties given below assume $y \in S_i(x)$. Note that this simply means that y is proposed from x with positive probability.

Standard Gibbs Sampler. The standard Gibbs sampler is retrieved by taking $H_1 = H$, $H_2 \equiv 0$, $\Omega = \Lambda^L$, and $S_i(x) = \{y \in \Omega: y_{(a_i)} = x_{(a_i)}\}$. It is easily verified that if $y \in S_i(x)$, the test ratio is one, so $P_T^{(i)} = Q_T^{(i)}$. Proposals are always accepted.

Multiple-site Updates. In the extension given in Geman and Geman [4], the site visitation schedule is $(A_i)_{i=0,1,2,\dots}$, where $A_i \subset L$. The values of the field on a subset, not necessarily a singleton, are updated at each time i . In the context presented here, $H_1 = H$

and $S_i(x) = \{y \in \Lambda^L: y_{(A_i)} = x_{(A_i)}\}$. One expects that as $|A_i| \rightarrow |L|$, transitions $P_T^{(i)}(x, \bullet)$ become “close” to Π_T ; indeed, $A_i \equiv L$ corresponds to direct sampling, in which case $P_T^{(i)}(x, \bullet) = \Pi_T$. This closeness is made precise in the discussion on optimality later in this chapter.

Hybrid Gibbs Sampler/Metropolis Algorithm. Let $S_i(x) = \{y \in \Lambda^L: y_{(a_i)} = x_{(a_i)}\}$, as in the standard Gibbs sampler, but let $H = H_1 + H_2$. In this case the test ratio does not involve H_1 whenever $y \in S_i(x)$. Here, for example, H_1 can be either the prior or data term in a Bayesian formulation. A new state is proposed by sampling from a one-term Gibbs distribution locally as in the standard Gibbs sampler. The acceptance decision, as in the Metropolis algorithm, is based on a test ratio, but here involving only the second energy term. The advantage of the hybrid over the Metropolis algorithm is that the state proposal is consistent with at least one of the energy terms. The advantage over the Gibbs sampler is that it may be simple to sample from the one-term conditional distribution, as when the energy is quadratic, but difficult to sample from the two-term conditional distribution.

In the following, take $H_1 = H$, let (a_i) be a site visitation schedule, and let $S_i(x) = \{y \in \Omega: y_{(a_i)} = x_{(a_i)}, y_{a_i} \in E_{a_i}(x)\}$, where $E_s(x) \subset \{u: (u, x_{(s)}) \in \Omega\}$. Note that if E_s does not depend on x_s the test ratio is always one whenever $y \in S_i(x)$. The standard Gibbs sampler has $E_s(x) \equiv \Lambda$ and Ω a product space.

Restricted Image Spaces. An example in which the sample space is *not* a product space, i.e., $E_s(x)$ actually depends on x , is given in Geman *et al.* [3] and Yang [15]. There E_s is chosen in a manner that is both sensible in an image reconstruction setting and allows a reduction in computation. The choice has the very special properties that the test ratio is always one when $y \in S_i(x)$ (because $E_s(x)$ does not depend on x_s), and $E_s(x) = \{u: (u, x_{(s)}) \in \Omega\}$.

Edge Dependence. This example is also relevant to image reconstruction. Let $\Omega = \Lambda^L$ and $E_s(x) = [x_s - g(x_{(s)}), x_s + g(x_{(s)})] \cap \Lambda$; i.e., at site s a range of values about the current state is considered where the range may depend on the current state at all sites except s . A reasonable choice for the function g is $g(x_{(s)}) = \max(\Delta, \max_{t \in \eta_s} x_t - \min_{t \in \eta_s} x_t)$, where η_s is some neighborhood of s and $\Delta \geq 1$. In particular, where there is an “edge” or boundary in the image, a larger range of values is considered than where the scene is relatively constant. This has the very desirable property of naturally concentrating the computation around the edges, which was one motivation for the work on restricted image spaces.

Back to Theorem 5.1

Before proving Theorem 5.1, some preliminary results are established.

Lemma 5.1. The reversibility condition $P_T^{(i)}(x, y)\Pi_T(x) = P_T^{(i)}(y, x)\Pi_T(y)$ is satisfied by Π_T and $P_T^{(i)}$. In particular, the invariance condition holds; i.e., $\sum_{x \in \Omega} P_T^{(i)}(x, y)\Pi_T(x) = \Pi_T(y)$.

Proof. The case $y = x$ is trivial. Let $y \neq x$, then

$$\begin{aligned} P_T^{(i)}(x, y)\Pi_T(x) &= Q_T^{(i)}(x, y)\Pi_T(x) \min(1, \rho_T^{(i)}(x, y)) \\ &= \min(Q_T^{(i)}(x, y)\Pi_T(x), Q_T^{(i)}(x, y)\Pi_T(x)\rho_T^{(i)}(x, y)) \end{aligned} \quad (5.9)$$

Further,

$$\begin{aligned} Q_T^{(i)}(x, y)\Pi_T(x) &= \frac{\exp(-(H_1(x) + H_2(x) + H_1(y))/T)}{Z_T \sum_{z \in \mathcal{S}_i(x)} \exp(-H_1(z)/T)} 1_{\mathcal{S}_i(x)}(y) \\ Q_T^{(i)}(x, y)\Pi_T(x)\rho_T^{(i)}(x, y) &= \frac{\exp(-(H_1(y) + H_2(y) + H_1(x))/T)}{Z_T \sum_{z \in \mathcal{S}_i(y)} \exp(-H_1(z)/T)} 1_{\mathcal{S}_i(x)}(y) \\ &= Q_T^{(i)}(y, x)\Pi_T(y) \end{aligned} \quad (5.10)$$

(Recall that $1_{\mathcal{S}_i(x)}(y) = 1_{\mathcal{S}_i(y)}(x)$). Hence, with $g(x, y) = Q_T^{(i)}(x, y)\Pi_T(x)$,

$$P_T^{(i)}(x, y) \Pi_T(x) = \min(g(x, y), g(y, x)) = P_T^{(i)}(y, x) \Pi_T(y) \quad (5.11)$$

which completes the proof.

Lemma 5.2. $P_T^{(i)}(x, y) \geq \frac{1}{|\Omega|} \exp(-c/T)$ for $y \in S_i(x)$, where

$$c = \max H_1 - \min H_1 + \max H_2 - \min H_2$$

Proof. Let $y \neq x$, then

$$Q_T^{(i)}(x, y) = \frac{\exp(-(H_1(y) - \min H_1)/T)}{\sum_{z \in S_i(x)} \exp(-(H_1(z) - \min H_1)/T)} \geq \frac{\exp(-(\max H_1 - \min H_1)/T)}{\sum_{z \in S_i(x)} \exp(-(H_1(z) - \min H_1)/T)} \quad (5.12)$$

$$\rho_T^{(i)}(x, y) \geq \frac{\exp(-(\max H_2 - \min H_2)/T) \sum_{z \in S_i(x)} \exp(-(H_1(z) - \min H_1)/T)}{\sum_{z \in S_i(y)} \exp(-(H_1(z) - \min H_1)/T)} \quad (5.13)$$

$$\geq \frac{1}{|\Omega|} \exp(-(\max H_2 - \min H_2)/T) \sum_{z \in S_i(x)} \exp(-(H_1(z) - \min H_1)/T)$$

Hence $P_T^{(i)}(x, y) = Q_T^{(i)}(x, y) \min(1, \rho_T^{(i)}(x, y)) \geq \frac{1}{|\Omega|} \exp(-c/T)$. Let $y = x$, then

$$P_T^{(i)}(x, y) \geq Q_T^{(i)}(x, y) \geq \frac{1}{|\Omega|} \exp(-(\max H_1 - \min H_1)/T) \geq \frac{1}{|\Omega|} \exp(-c/T) \quad (5.14)$$

Lemma 5.3. For $x \in \Omega$, $\sum_{k=0}^{\infty} |\Pi_{T_{k+1}}(x) - \Pi_{T_k}(x)| < \infty$.

Proof. Let $\Pi_{T_k}(x) = \frac{p_k}{q_k}$, where

$$0 \leq p_k = \exp(-(H(x) - \min H)/T_k) \leq 1 \quad (5.15)$$

$$1 \leq q_k = \sum_{x \in \Omega} \exp(-(H(x) - \min H)/T_k) \leq |\Omega| \quad (5.16)$$

with $H = H_1 + H_2$. Further,

$$\begin{aligned}
|\Pi_{T_{k+1}}(x) - \Pi_{T_k}(x)| &= \left| \frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k} \right| = \frac{1}{q_k q_{k+1}} |p_k q_{k+1} - p_{k+1} q_k| \leq |p_k q_{k+1} - p_{k+1} q_k| \\
&\leq p_k |q_{k+1} - q_k| + q_k |p_k - p_{k+1}| \leq |q_{k+1} - q_k| + |\Omega| |p_k - p_{k+1}|
\end{aligned} \tag{5.17}$$

Since (T_k) is eventually decreasing, for some k_0 , $p_{k+1} \leq p_k$, $q_{k+1} \leq q_k$, for all $k \geq k_0$, and

$$\begin{aligned}
&\sum_{k=k_0}^{\infty} |\Pi_{T_{k+1}}(x) - \Pi_{T_k}(x)| \\
&\leq \sum_{k=k_0}^{\infty} q_k - q_{k+1} + |\Omega| \sum_{k=k_0}^{\infty} p_k - p_{k+1} = q_{k_0} - \lim q_k + |\Omega| (p_{k_0} - \lim p_k) < \infty
\end{aligned} \tag{5.18}$$

Proof of Theorem 5.1. The method of proof is that in Geman [2]. By Isaacson and Madsen [10, Theorem V.4.3], the results of Lemmas 5.1 and 5.3 are sufficient if $(X_i)_{i=0,1,2,\dots}$ is weakly ergodic. Weak ergodicity is implied by $\sum_{k=0}^{\infty} \alpha(P_{T_k}^{(\tau_k, \tau_{k+1})}) = \infty$,

where $P_{T_k}^{(\tau_k, \tau_{k+1})} = \prod_{i=\tau_k}^{\tau_{k+1}-1} P_{T_{k_i}}^{(i)} = \prod_{i=\tau_k}^{\tau_{k+1}-1} P_{T_k}^{(i)}$ and α denotes the ergodic coefficient. By

hypothesis, given k and $x, y \in \Omega$ there exists a sequence

$x = x_{\tau_k}, x_{\tau_k+1}, \dots, x_{\tau_{k+1}-1}, x_{\tau_{k+1}} = y$ such that $x_{i+1} \in S_i(x_i)$ for $i = \tau_k, \tau_k + 1, \dots, \tau_{k+1} - 1$.

So, using Lemma 5.2,

$$\begin{aligned}
P_{T_k}^{(\tau_k, \tau_{k+1})}(x, y) &= \left(\prod_{i=\tau_k}^{\tau_{k+1}-1} P_{T_k}^{(i)} \right)(x, y) \geq \prod_{i=\tau_k}^{\tau_{k+1}-1} P_{T_k}^{(i)}(x_i, x_{i+1}) \geq \prod_{i=\tau_k}^{\tau_{k+1}-1} \frac{1}{|\Omega|} \exp(-c/T_k) \\
&= \left(\frac{1}{|\Omega|} \exp(-c/T_k) \right)^{\tau_{k+1}-\tau_k} \geq \frac{\exp(-nc/T_k)}{|\Omega|^n}
\end{aligned} \tag{5.19}$$

Now, for $\liminf_{k \rightarrow \infty} T_k \log k \geq nc$

$$\begin{aligned}
\sum_{k=0}^{\infty} \alpha(P_{T_k}^{(\tau_k, \tau_{k+1})}) &= \sum_{k=0}^{\infty} \min_{x, y \in \Omega} \sum_{z \in \Omega} \min(P_{T_k}^{(\tau_k, \tau_{k+1})}(x, z), P_{T_k}^{(\tau_k, \tau_{k+1})}(y, z)) \\
&\geq \sum_{k=0}^{\infty} \min_{x, y \in \Omega} \sum_{z \in \Omega} \frac{\exp(-nc/T_k)}{|\Omega|^n} = \frac{1}{|\Omega|^{n-1}} \sum_{k=0}^{\infty} \exp(-nc/T_k) = \infty
\end{aligned} \tag{5.20}$$

and the proof is complete.

Note. It is not necessary that temperature be constant on the intervals $\{[\tau_k, \tau_{k+1}), k \geq 0\}$, i.e., that the schedule be of the form (T_{k_i}) . Any eventually decreasing sequence (T'_i) which satisfies $T'_i \geq T_{k_i}$ for all i will suffice.

Extension

The level of generality of the above was chosen to adhere to the basic philosophy of the Gibbs sampler, and to allow practically useful sampling schemes without imposing undue complication. With respect to the design of reversible transition operators, a further generalization of the above is suggested by a non-stationary version of the Hastings extension to the Metropolis algorithm in its full generality. Given a distribution $\Pi^{(i)}$, transition operators which satisfy the reversibility condition, and therefore invariance, are given by

$$P^{(i)}(x,y) = \begin{cases} Q^{(i)}(x,y)\alpha^{(i)}(x,y) & Q^{(i)}(x,y) > 0, y \neq x \\ 1 - \frac{\sum_{Q^{(i)}(x,z) > 0, z \neq x} Q^{(i)}(x,z)\alpha^{(i)}(x,z)}{Q^{(i)}(x,x)} & y = x \\ 0 & \text{otherwise} \end{cases} \quad (5.21)$$

where $Q^{(i)}$ is arbitrary except that $Q^{(i)}(x,y) > 0 \Leftrightarrow Q^{(i)}(y,x) > 0$ must be satisfied¹ for all x,y and

$$\alpha^{(i)}(x,y) = \frac{s^{(i)}(x,y)}{1 + \frac{\Pi^{(i)}(x) Q^{(i)}(x,y)}{\Pi^{(i)}(y) Q^{(i)}(y,x)}} \quad (5.22)$$

with $s^{(i)}$ symmetric and chosen so that $0 \leq \alpha^{(i)} \leq 1$. This framework gives a propose/decide scheme; with the current state x , y is proposed with probability $Q_T^{(i)}(x,y)$,

¹This condition is left out of the Hastings paper, but there is a divide-by-zero problem without it.

and accepted with probability $\alpha^{(i)}(x, y)$. For an optimality result on the Hastings acceptance schemes, see Peskun [13].

Optimality

The extended Gibbs sampler enjoys some optimality properties which are treated below. Let Π_T be a probability mass function of the form (5.5), and let (S_i) be a sequence of maps from Ω to subsets of Ω as above. Define the proposal distribution

$$\Pi_T^{(1)}(x) = \frac{\exp(-H_1(x)/T)}{\sum_{y \in \Omega} \exp(-H_1(y)/T)}$$

and define $\Phi_{i,x}$ to be the set of probability mass functions f which satisfy $f(y) = 0$ for all $y \notin S_i(x)$. Also, let

$$I(f, g) = \sum_y f(y) \log(f(y)/g(y)) \quad (5.23)$$

denote the relative entropy between two probability mass functions f and g . Finally, define $R(Q_T^{(i)}, \Pi_T)$ to be the set of time-reversible propose/decide transition operators associated with $Q_T^{(i)}$ and Π_T :

$$\begin{aligned} R(Q_T^{(i)}, \Pi_T) = \{r: \Omega^2 \rightarrow [0, 1]: \sum_{y \in \Omega} r(x, y) = 1 \forall x, \Pi_T(x)r(x, y) = \Pi_T(y)r(y, x) \forall (x, y), \\ r(x, y) \leq Q_T^{(i)}(x, y) \forall (x, y): y \neq x\} \end{aligned} \quad (5.24)$$

The last condition ensures that the operators have the form (5.21) with $0 \leq \alpha^{(i)} \leq 1$.

Theorem 5.2. $Q_T^{(i)}$ is optimal in the following sense. For each i and all $x \in \Omega$,

$$Q_T^{(i)}(x, \bullet) = \arg \min_{f \in \Phi_{i,x}} I(f, \Pi_T^{(1)}) \quad (5.25)$$

Further, $P_T^{(i)}$ is optimal in the following sense.

$$P_T^{(i)} = \arg \max_{r \in R(Q_T^{(i)}, \Pi_T)} \#\{(x, y) \in \Omega^2 : r(x, y) = Q_T^{(i)}(x, y)\} \quad (5.26)$$

In words, $P_T^{(i)}$ is the time-reversible transition operator which matches $Q_T^{(i)}$ most often. If $H_2 \equiv 0$, then $Q_T^{(i)}$ is in a sense “closest” to Π_T . For the case of a posterior distribution when $H_2 \neq 0$, one of three situations is standard: 1) H_1 is the prior energy, and $Q_T^{(i)}$ is closest to the prior, which may be viewed as the joint distribution of the scene and data with the data “averaged out”; 2) H_1 is the data term, and $Q_T^{(i)}$ is closest to the distribution obtained by normalizing the likelihood function; 3) For the Metropolis case where $H_1 \equiv 0$, $Q_T^{(i)}$ is closest to the uniform distribution on Ω .

The following lemma is preliminary to the proof.

Lemma 5.4. Let g be a probability mass function on Ω and define $g_S = \frac{g}{\sum_{y \in S} g(y)} 1_S$ for a

subset $S \subset \Omega$. Define Φ_S to be the set of probability mass functions f which satisfy $f(y) = 0$ for all $y \notin S$. Then

$$g_S = \arg \min_{f \in \Phi_S} I(f, g) \quad (5.27)$$

Proof.

$$\begin{aligned} \min_{f \in \Phi_S} I(f, g) &= \min_f \sum_{y \in S} f(y) \log(f(y)/g(y)) \\ &= \min_f \sum_{y \in S} f(y) \left(\log\left(\frac{f(y)}{g_S(y)}\right) - \log\left(\sum_{z \in S} g(z)\right) \right) \\ &= \min_f \sum_{y \in S} f(y) \log\left(\frac{f(y)}{g_S(y)}\right) + \log\left(\frac{1}{\sum_{z \in S} g(z)}\right) \\ &= \min_f I(f, g_S) + \log\left(\frac{1}{\sum_{z \in S} g(z)}\right) \end{aligned} \quad (5.28)$$

It is a standard property of relative entropy that it is greater than or equal to zero with equality if and only if its two arguments are equal (Kullback [12]). Hence $f = g_S$ is

optimum for the first term on the right-hand side above, in which case

$I(g_S, g) = \log\left(\frac{1}{\sum_{y \in S} g(y)}\right)$. Thus, as is intuitively plausible, the two distributions are close

when $\sum_{y \in S} g(y)$ is close to one, i.e., g has a lot of mass on S .

Note. Clearly, as $|S| \rightarrow |\Omega|$, $\log\left(\frac{1}{\sum_{y \in S} g(y)}\right) \rightarrow 0$. In the multiple-site update Gibbs sampler

presented earlier, as $|A_i| \rightarrow |L|$, $|S_i(x)| \rightarrow |\Omega|$. As a result, the more sites that are simultaneously updated, the closer $P_T^{(i)}(x, \bullet)$ gets to Π_T in the sense of relative entropy; however, “close” is somewhat imprecise as relative entropy is not a metric.

Proof of Theorem 5.2. The proof of the first result of Theorem 5.2 is an application of Lemma 5.4. To prove the second, let $r \in R(Q_T^{(i)}, \Pi_T)$. By the reversibility condition

$r(y, x) = \Pi_T(x)r(x, y)/\Pi_T(y)$; by the definition of $\rho_T^{(i)}$,

$Q_T^{(i)}(y, x)\rho_T^{(i)}(y, x) = \Pi_T(x)Q_T^{(i)}(x, y)/\Pi_T(y)$. Consequently,

$$r(x, y) \leq Q_T^{(i)}(x, y) \Leftrightarrow r(y, x) \leq Q_T^{(i)}(y, x)\rho_T^{(i)}(y, x) \quad \forall (x, y) \quad (5.29)$$

A property of r , by definition of $R(Q_T^{(i)}, \Pi_T)$ given in (5.26), is that for (x, y) such that $y \neq x$, $r(x, y) \leq Q_T^{(i)}(x, y)$, so with the above (5.29), $r(x, y) \leq Q_T^{(i)}(x, y) \min(1, \rho_T^{(i)}(x, y))$ $\forall (x, y): y \neq x$. Now,

$$\begin{aligned} \#\{(x, y): r(x, y) = Q_T^{(i)}(x, y)\} \\ = \#\{(x, y): y \neq x, r(x, y) = Q_T^{(i)}(x, y)\} + \#\{(x, x): x \in \Omega, r(x, x) = Q_T^{(i)}(x, x)\} \end{aligned} \quad (5.30)$$

Bounding the first term,

$$\begin{aligned} \#\{(x, y): y \neq x, r(x, y) = Q_T^{(i)}(x, y)\} \\ = \#\{(x, y): y \neq x, r(x, y) = Q_T^{(i)}(x, y) \min(1, \rho_T^{(i)}(x, y)), \rho_T^{(i)}(x, y) \geq 1\} \\ \leq \#\{(x, y): y \neq x, \rho_T^{(i)}(x, y) \geq 1\} \end{aligned} \quad (5.31)$$

and since $\rho_T^{(i)}(x, y) = 1/\rho_T^{(i)}(y, x)$,

$$= \#\{(x,y):y \neq x, \rho_T^{(i)}(x,y) \neq 1\}/2 + \#\{(x,y):y \neq x, \rho_T^{(i)}(x,y) = 1\} \quad (5.32)$$

To bound the second term, for $x \in \Omega$, the three conditions $r(x,y) \leq Q_T^{(i)}(x,y) \forall y \neq x$, $\sum_y r(x,y) = 1$, and $\sum_y Q_T^{(i)}(x,y) = 1$ imply that $r(x,x) \geq Q_T^{(i)}(x,x)$, with equality iff

$$\begin{aligned} r(x,y) &= Q_T^{(i)}(x,y) \quad \forall y \neq x \\ \Leftrightarrow r(x,y) &= Q_T^{(i)}(x,y) \min(1, \rho_T^{(i)}(x,y)) \quad \forall y \neq x, \rho_T^{(i)}(x,y) \geq 1 \quad \forall y \neq x \end{aligned} \quad (5.33)$$

Hence,

$$\begin{aligned} &\#\{(x,x):x \in \Omega, r(x,x) = Q_T^{(i)}(x,x)\} \\ &= \#\{x:r(x,y) = Q_T^{(i)}(x,y) \min(1, \rho_T^{(i)}(x,y)) \quad \forall y \neq x, \rho_T^{(i)}(x,y) \geq 1 \quad \forall y \neq x\} \\ &\leq \#\{x:\rho_T^{(i)}(x,y) \geq 1 \quad \forall y \neq x\} \end{aligned} \quad (5.34)$$

The upper bounds for both terms are achieved when

$$r(x,y) = Q_T^{(i)}(x,y) \min(1, \rho_T^{(i)}(x,y)) \quad \forall (x,y):y \neq x \quad (5.35)$$

This completes the proof.

For some related work based on relative entropy, see Goutsias [6], [7].

Construction

The proof suggests the most general construction for elements of $R(Q_T^{(i)}, \Pi_T)$, the time-reversible, propose/decide operators associated with a given Π_T and $Q_T^{(i)}$. Let $A \subset \Omega^2$ satisfy $(x,x) \notin A \quad \forall x \in \Omega$ and $(x,y) \in A \Leftrightarrow (y,x) \in A \quad \forall (x,y) \in \Omega^2$. For each $(x,y) \in A$ let $r(x,y)$ satisfy $r(x,y) \leq Q_T^{(i)}(x,y) \min(1, \rho_T^{(i)}(x,y))$, and set $r(y,x) = \Pi_T(x)r(x,y)/\Pi_T(y)$. Finally for each $x \in \Omega$ let $r(x,x) = 1 - \sum_{y \neq x} r(x,y)$.

References

1. J. Besag and P. Green (1993). "Spatial statistics and Bayesian computation," *J. Royal Statist. Soc. B*, 55, pp. 25-37 (preprint).
2. D. Geman, (1991). "Random fields and inverse problems in imaging," *Lecture Notes in Mathematics*, Vol. 1427, Springer-Verlag.
3. D. Geman, G. Reynolds, and C. Yang, "Stochastic algorithms for restricted image spaces and deblurring," to be published in *Markov Random Fields: Theory and Applications*. New York: Academic.
4. D. Geman and S. Geman, (1987). "Relaxation and annealing with constraints," *Complex Systems Technical Report No. 35*, Div. of Applied Mathematics, Brown University.
5. S. Geman and D. Geman, (1984). "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, 6, pp. 721-741.
6. J. K. Goutsias, (1991). "A theoretical analysis of Monte Carlo algorithms for the simulation of Gibbs random field images," *IEEE Trans. Inform. Theory*, 37, pp. 1618-1628.
7. J. K. Goutsias, (1993). Author's reply to "Comments on 'A theoretical analysis of Monte Carlo algorithms for the simulation of Gibbs random field images,'" *IEEE Trans. Inform. Theory*, 39, pp. 310-311.
8. P. J. Green and X. Han, (1992). "Metropolis methods, Gaussian proposals, and antithetic variables," preprint.
9. W. K. Hastings, (1970). "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57, pp. 97-109.
10. D. Isaacson and R. Madsen, (1976). *Markov Chains: Theory and Applications*. New York: John Wiley and Sons.
11. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, (1953). "Equations of state calculations by fast computing machines," *J. Chemical Physics*, 21, pp. 1087-1091.
12. S. Kullback, (1959). *Information Theory and Statistics*. New York: John Wiley and Sons.
13. P. H. Peskun (1973). "Optimum Monte-Carlo sampling using Markov chains," *Biometrika*, 60, pp. 607-612.
14. L. Tierney (1991). "Exploring posterior distributions using Markov chains," Proceedings of the 1991 Interface meeting (preprint).

15. C. Yang, (1991). "Stochastic methods for image restoration," Ph.D. thesis, Dept. Math. Stat., Univ. Mass., Amherst, MA.

CHAPTER 6

RADARCLINOMETRY ALGORITHM

This chapter provides a computational procedure for deriving surface shape from a radar image of the surface. A sampling algorithm is described based on the extended Gibbs sampler, and a hierarchical, multi-resolution technique suggested by the self-similarity result for the stabilizing Gibbs distribution is described. The chapter also includes experimental results of an application of the method to radar images of the surface of Venus acquired by the Magellan probe.

Posterior Distribution

The point estimate employed in the radarclinometry algorithm is the posterior mean $E(X|U = u)$. This choice provides good results, and does not require selection of an *ad hoc* temperature schedule as in the case of a posterior mode. Distributions do not seem quite “peaked” enough to warrant the use of a sample from the posterior distribution as a point estimate.

As given in Chapter 3 with the prior energy specified in Chapter 4, the posterior distribution is $P(X \in dx|U = u) = \text{const.} \exp(-H_u(x))dx$, where

$$H_u(x) = \left\| T_{(2,\sigma,\rho)} x \right\|^2 + \lambda \sum_{s \in F} h(\gamma \varphi_s(x), u_s) \quad (6.1)$$

and $h(a, b) = b/a - \log(b/a) - 1$. The set F is the domain of the radar image, as in Chapter 2.

Let L be the lattice on which the surface is defined. For computational purposes the continuum state space \mathbf{R}^L for x is approximated by a finite state space Λ^L , where Λ is an appropriate subset of the real numbers. Let

$$\pi(x|u) = \frac{\exp(-H_u(x))}{\sum_{x \in \Lambda^L} \exp(-H_u(x))} \quad (6.2)$$

for $x \in \Lambda^L$. The approximation (due to restriction to a finite sample space) to the posterior mean is $\sum_{x \in \Lambda^L} x \pi(x|u)$.

Since $\#(\Lambda^L) = \#\Lambda^{\#L}$, direct computation of the expectation is computationally prohibitive even for $\#\Lambda = 2$ if the image lattice has a practically useful size. For this reason a Monte Carlo estimate of the expectation is used, given by the average of a sequence x_i consisting of samples from $\pi(x|u)$. However, there is the remaining difficulty of sampling from $\pi(x|u)$, which is addressed in the next section.

Sampling Algorithm

The sampling algorithm is a version of the extended Gibbs sampler given in Chapter 5. In particular, $H_1(x) = \|T_{(2,\sigma,\rho)}x\|^2$, the quadratic prior term, $H_2(x) = \lambda \sum_{s \in F} h(\gamma\phi_s(x), u_s)$, the data term, and $T_k \equiv 1$. Single-site visitation is used as in the standard Gibbs sampler, where the site visitation schedule is chosen so that temporally neighboring sites (sites visited at contiguous instants) are spatially distant. This helps to prevent artifacts which result from standard orderings of the lattice.

The motivation for the sampling algorithm, indeed the extension of the Gibbs sampler, is as follows. The standard choices for Monte Carlo sampling are the standard Gibbs sampler and the Metropolis algorithm. The standard Gibbs sampler is computationally unwieldy because the local state space is large. It has $\#\Lambda$ elements, and for each one the image formation operation, a complex calculation (see (3.5) and (5.3)), must be carried out in calculating the data term. The Metropolis algorithm involves only two image formation operations (one in the numerator of the test ratio, one in the denominator), but the state proposal scheme is *ad hoc*. Because the prior is Gaussian the

local distribution based solely on the prior is *univariate* Gaussian, from which a sample is easy to generate on a computer. Therefore, an attractive scheme is to propose states by sampling the local prior, and to accept based on a Metropolis-type ratio involving only the data term. In this way the proposal, easily generated, is at least consistent with the prior, and the Metropolis acceptance decision requires only two image formation operations. This scheme is precisely the extended Gibbs sampler described earlier.

The prescription results in the following algorithm. Assume that site (i, j) is being updated, and define $x_{[i,j]} = (x_{(i',j')})_{(i',j') \neq (i,j)}$, i.e., a vector containing all elements of x except for x_{ij} . A new state $(v, x_{[i,j]})$ is proposed, where v is drawn from the univariate normal distribution

$$\pi_1^{(i,j)}(v) = c(x_{[i,j]}) \exp\left(-\|T_{(2,\sigma,\rho)}(v, x_{[i,j]})\|^2\right) \quad (6.3)$$

where $c(x_{[i,j]})$ is the normalizing constant. The local conditional mean and variance is obtained by completing the square in the exponent. The mean is

$$\begin{aligned} \mu_{ij} = & a_1(x_{i-1,j} + x_{i+1,j}) + a_2(x_{i,j-1} + x_{i,j+1}) \\ & - a_3(x_{i-1,j-1} + x_{i+1,j-1} + x_{i-1,j+1} + x_{i+1,j+1}) \\ & - a_4(x_{i-2,j} + x_{i+2,j}) - a_5(x_{i,j-2} + x_{i,j+2}) \end{aligned} \quad (6.4)$$

where

$$\begin{aligned} a_4 = 1/\left(6(1+\rho^4) + 8\rho^2\right) & \quad a_5 = a_4\rho^4 \\ a_1 = 4a_4(1+\rho^2) & \quad a_2 = a_1\rho^2 \quad a_3 = 2a_4\rho^4 \end{aligned}$$

and the variance is $a_4\sigma^2$. Spatial isotropy of the surface is assumed, so $\rho = n_2/n_1$.

The proposal is accepted with probability

$$\pi_2^{(i,j)}(v) = \min\left(1, \exp\left(-\lambda \sum_{s \in F} h(\gamma\phi_s(v, x_{[i,j]}), u_s)\right) / \exp\left(-\lambda \sum_{s \in F} h(\gamma\phi_s(x), u_s)\right)\right) \quad (6.5)$$

Because of cancellation between the numerator and denominator, the sums can be taken only over the set $F_{ij} = \{s: \varphi_s(v, x_{[i,j]}) \neq \varphi_s(x)\}$, which is a relatively small subset of F if the point spread function (approximated by truncating) has relatively local support. This together with the fact that sampling from the local prior at (i, j) involves only 12 elements of x implies that the posterior distribution has a relatively local neighborhood structure, an important computational consideration.

Implementation notes. In proposing changes based on the local prior, only the local mean changes from site to site; the local variance is homogeneous. The quadratic associated with the normal distribution is never evaluated, as would be the case with the standard Gibbs sampler or the Metropolis algorithm. The Gaussian random variate is generated using the inverse c.d.f. lookup table method. In updating a site, the time for the image-formation calculation for the current state can be reduced by storing the data-term computations carried out at the previous site visit (when either the current or proposed energy was the now-current energy). In this way the Metropolis acceptance scheme requires that only the data term for the *proposed* state need be evaluated at each site update, reducing to one the number of image formation operations.

Multi-resolution Scheme

In theory (Theorem 5.1), given an arbitrarily large amount of time to visit sites and carry out local computations, the extended Gibbs sampler generates samples from the global posterior distribution, which can be used in Monte Carlo estimation. In practice a limited amount of time is available. This is the motivation for the multi-resolution scheme adopted in sampling from the posterior distribution. The basic idea is prevalent in image processing (e.g., Gidas [2], Rosenfeld [3]).

A multi-resolution hierarchy of surface estimations is carried out. At the top level the data and surface are defined with respect to the coarsest grids, and the grids are

successively finer at lower levels. The finest level is level 0. At time i , level l of the hierarchy, let the surface be denoted by $X^{(i,l)}$, with lattice size $(n_1^{(l)}, n_2^{(l)})$. Let $a_l = n_1^{(l-1)}/n_1^{(l)}$, and $b_l = n_2^{(l-1)}/n_2^{(l)}$, assumed to be positive integers. At each level the posterior mean is approximated by first sweeping the image, applying the Gibbs sampler until equilibrium of the Markov chain is assumed. Then realizations are averaged for a series of sweeps. The highest level in the hierarchy, level 3 in the experiments presented later, uses the reference surface as a starting point, which corresponds to $X^{(0,3)} \equiv 0$. In descending the hierarchy, the estimate at level l is upsampled by replication and used as the starting point for the next lower level in the hierarchy; i.e., if x denotes the final outcome for the Markov chain at level l , then the starting point at level $l-1$ is given by $X_{i,j}^{(0,l-1)} = x_{[i/a_l],[j/b_l]}$, for $(i,j) \in (n_1^{(l-1)}, n_2^{(l-1)})$.

The multi-resolution scheme is, practically speaking, the key to the success of the procedure, which is supported by experimental observations presented later. The scheme can be viewed as simply providing a good starting point for sweeping at the finest level of the hierarchy. The advantages of a multi-resolution method are intuitively clear. Information travels slowly across large distances when single site updates are used. The situation is rectified by allowing large-scale properties to be established early at high levels.

Model Consistency. A prominent difficulty in multi-resolution approaches is maintaining model consistency among resolution levels. Here this is accomplished in the following ways.

Prior. The self-similarity result given in Corollary 4.1 provides the appropriate relationship among the stabilizing Gibbs distribution parameter values. At level l let the parameter be denoted $(\sigma^{(l)}, \rho^{(l)})$. With $p = 2$, the parameter values satisfy

$$\begin{aligned}\sigma^{(l)} &= a_l^{-3/2} b_l^{1/2} \sigma^{(l-1)} \\ \rho^{(l)} &= (b_l/a_l) \rho^{(l-1)}\end{aligned}\tag{6.6}$$

With parameters so defined, the self-similarity result states that at a higher level the stabilizing Gibbs distribution is approximately equivalent to the marginal distribution of the field on the corresponding subset of sites at a lower level, i.e., $X_{a_i, b_{ij}}^{(l-1)} \approx X_{i,j}^{(l)}$ in distribution.

To demonstrate the effectiveness of the multi-resolution procedure, it was applied to a hierarchy of second order Gibbs distributions in standard form, which corresponds to the energy based solely on the prior. The result of both multi-resolution sampling and direct Gibbs sampling can be compared to a sample generated based on the representation theorem of Chapter 3. Fig. 6.1 shows the evolution of the multi-resolution sample, where a 16×16 grid is used at the highest level, a 128×128 grid at the lowest level. At each level 1024 sweeps of the lattice were carried out, with the highest-level starting point an image of zeros. The result should be compared for similarity in characteristics with Fig. 4.2, generated using the representation theorem. Demonstrating the advantage of the multi-resolution approach, Fig. 6.2 shows the result of applying the Gibbs sampler directly to a 128×128 lattice of zeros for 8192 sweeps. The result does not visually compare favorably with Fig. 4.2, even though more computation was required. This demonstrates that the starting point is crucial in using the Gibbs sampler, which motivates the multi-resolution scheme.

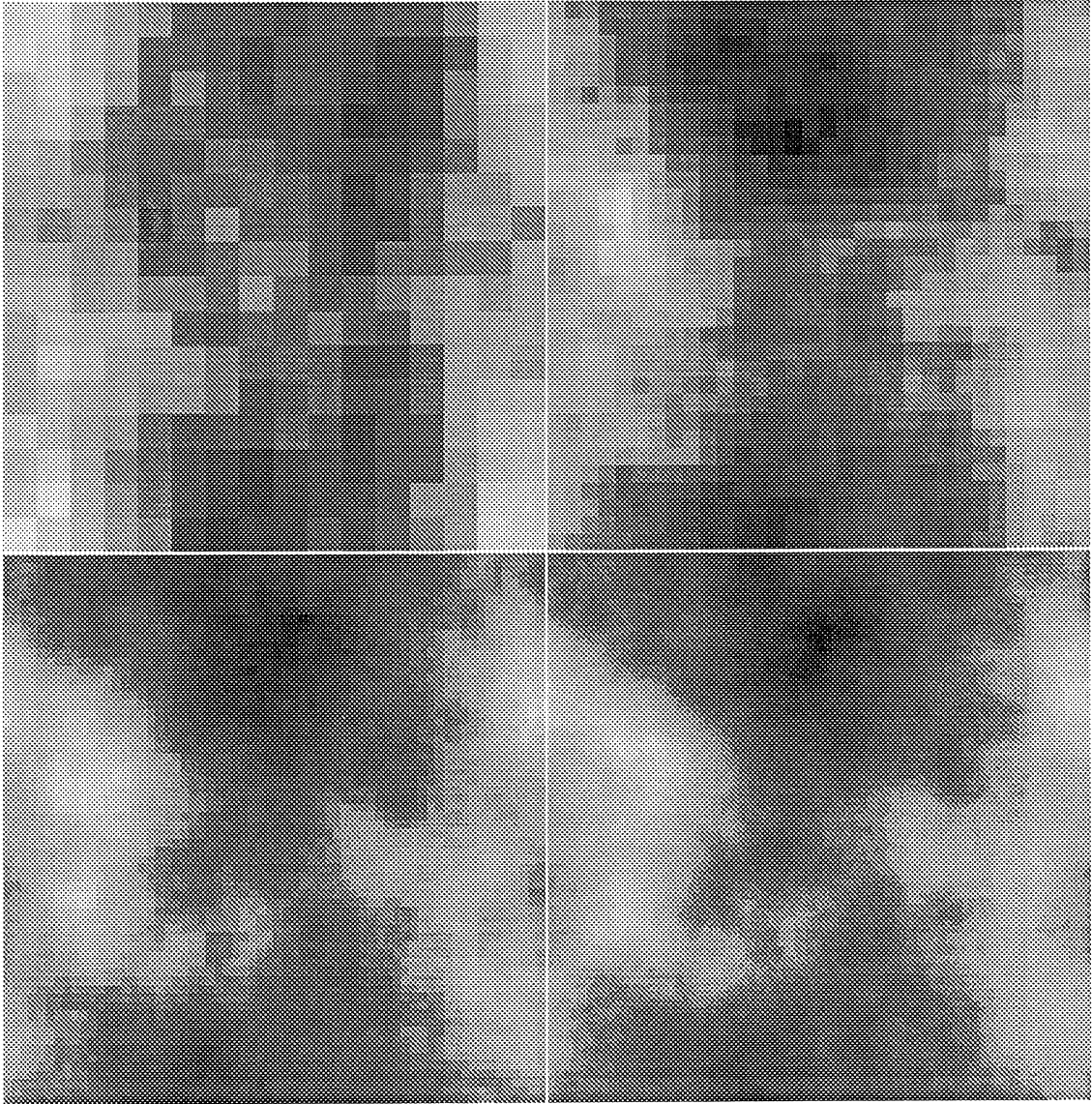


Fig. 6.1. Sample evolution.
top-left: 16×16 ; top-right: 32×32 ;
bottom-left: 64×64 ; bottom-right: 128×128 .

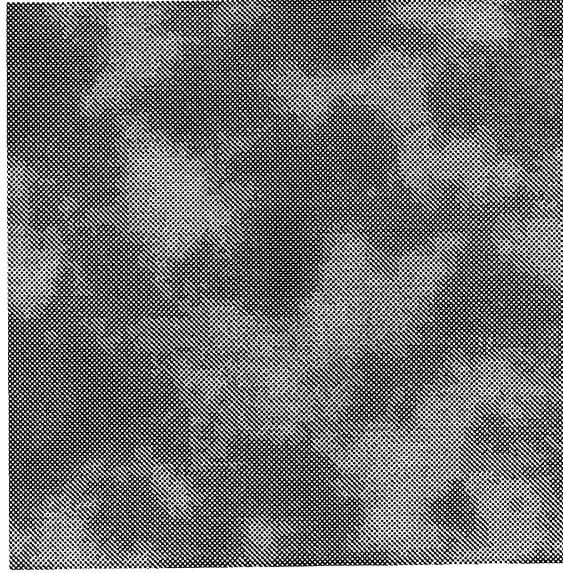


Fig. 6.2. Result of direct application of Gibbs sampler.

Data. The radar image data at level l are created from the data at level $l-1$ by smoothing with a product of Gaussian kernels (to prevent aliasing) and subsampling. That is, let $U^{(l)}$ denote the radar image at level l , where $U^{(0)}$ is the original data. Let $U^{(l)}$ have domain $F^{(l)}$, where the $F^{(l)}$ grow coarser with l , i.e., satisfy $F^{(l)} \subset F^{(l-1)}$. Then

$$U_{(r,y)}^{(l)} = c(r,y) \sum_{(r',y') \in S^{(l-1)}} \eta(r-r')\eta(y-y')U_{(r',y')}^{(l-1)}$$

for $(r,y) \in F^{(l)}$, where $\eta = N(0, t^{(l)})$, the mean-zero Gaussian probability density function with variance $t^{(l)}$, and $c(r,y) = \left(\sum_{(r',y') \in S^{(l-1)}} \eta(r-r')\eta(y-y') \right)^{-1}$ is a normalizing constant.

Thus the data hierarchy is created by ascending. The point spread function is smoothed and subsampled in the same way to maintain consistency with the data.

Experiments

The results of application of the method to synthetic and Magellan data are shown in this section. The reference surface is depicted in Fig. 3.2, i.e., is flat. In grey-scale depictions of image and surface data light areas correspond to large values, dark to small.

Synthetic Data. The first experiment presented is an application of the method to synthetic data, so that the effectiveness of the procedure can be evaluated in a situation with ground truth. A sample from the stabilizing Gibbs distribution of order two was obtained using the representation given in Chapter 4. From this image a subimage, shown in Fig. 6.3, was taken to avoid the effect of the periodic boundary conditions. The parameter ρ of the stabilizing Gibbs distribution was chosen so that the field is isotropic with respect to the surface grid. The scaling parameter σ was chosen to be typical of a mountainous natural terrain, by drawing a sample x from the $\sigma = 1$ distribution, and choosing the σ necessary to make the range of values for σx typical. In other words, σ is chosen to scale the sample so that the size of structures (mountains, valleys, etc.) is consistent with prior information in this regard. A noise-free radar image (Fig. 6.4) was then synthesized from the sample, to which 16-look speckle was added (Fig. 6.5). The sensor flight parameters used were taken to be the Magellan parameters (Table 6.1) of one of the later experiments.

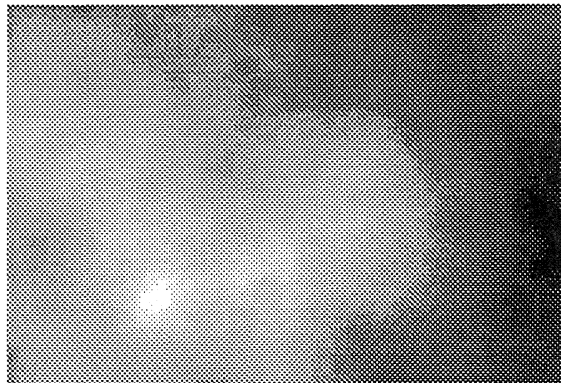


Fig. 6.3. Test surface.

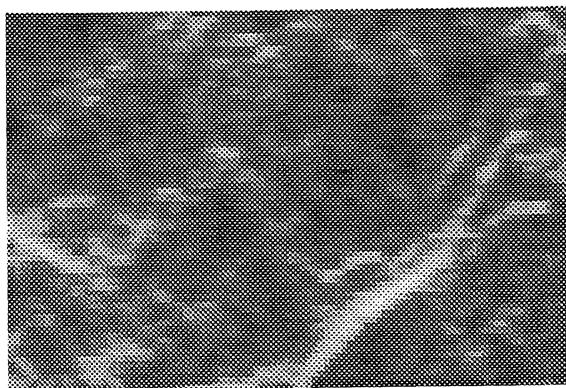


Fig. 6.4. Noise-free radar image.

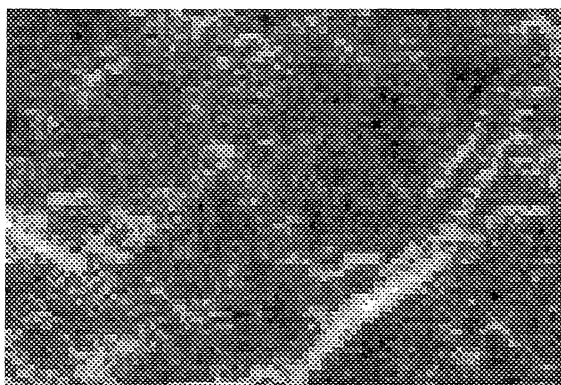


Fig. 6.5. 16-look radar image.

Table 6.1. Flight parameters.

incidence angle (θ_0)	23°
least range	5300 km
greatest range	5309.6 km
range resolution	230 m
azimuth resolution	120 m

Table 6.2 shows the remaining parameters used in the reconstruction procedure at each resolution level. Note that the grid size (n_1, n_2) was chosen so that synthetic radar images were free of discretization artifacts; this required a non-square lattice at resolution level 0.

Table 6.2. Reconstruction parameters.

level	n_1	n_2	F	σ	λ	iterations
0	128	256	96×64	0.14	16	32/32
1	64	64	48×32	0.2	32	256/128
2	32	32	24×16	0.4	64	256/128
3	16	16	12×8	0.8	128	512/256

An iteration is a complete update corresponding to the visitation of all sites. With respect to the iterations column in the table, the entries correspond to two cycles. As previously mentioned, during the first cycle the state distribution is being driven toward the invariant distribution. Iterations are averaged during the second cycle, forming a Monte Carlo estimate of the expectation. The evolution of the reconstruction is shown in Fig. 6.6, which shows the final surface state at each level. Figure 6.7 shows the noise-free radar image of the reconstruction, which should be compared to Fig. 6.4.

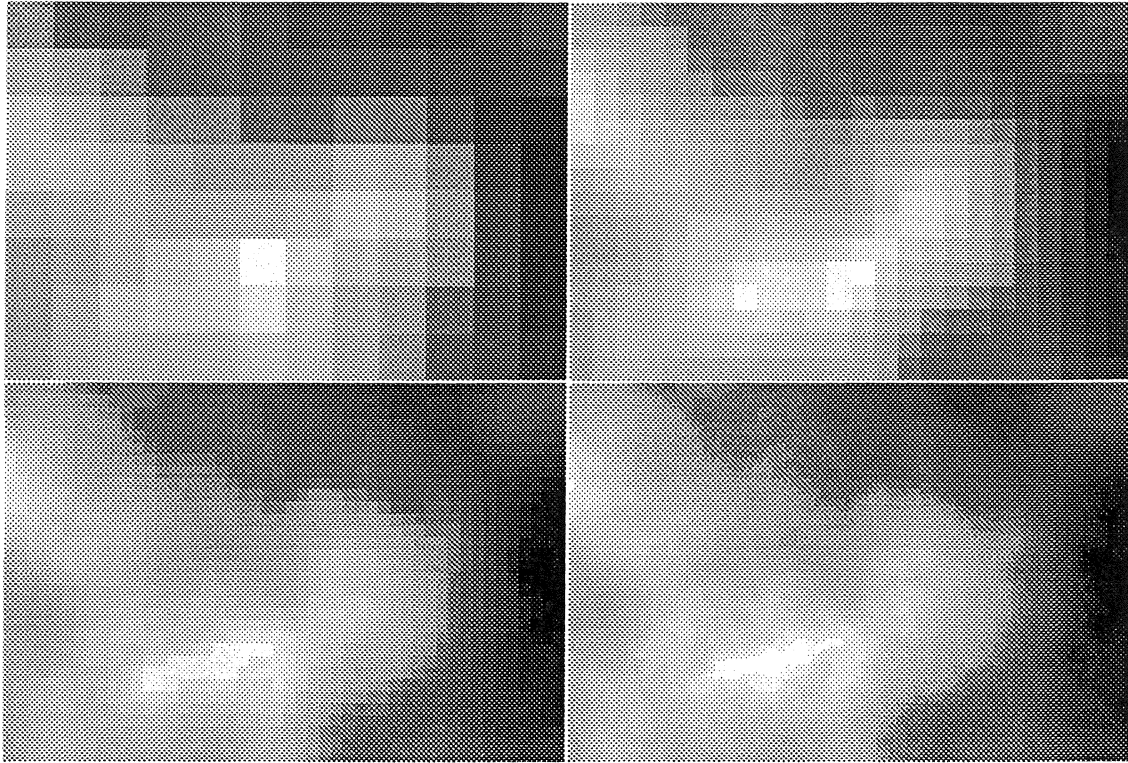


Fig. 6.6. Reconstructions; top-left: level 3; top-right: level 2; bottom-left: level 1; bottom-right: level 0.

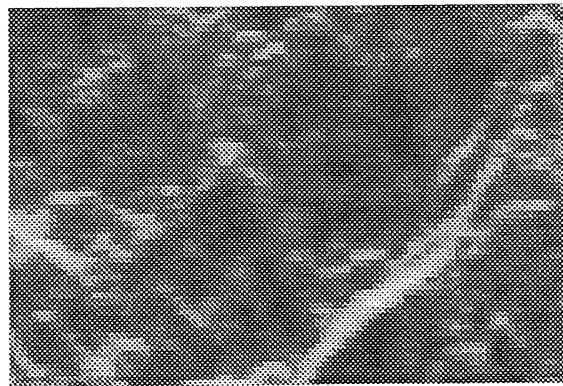


Fig. 6.7. Noise-free image of reconstruction.

Figure 6.8 shows evolution of the maximum likelihood estimate of the scale parameter σ , and Fig. 6.9 is the evolution of update ratio, i.e., the ratio of accepted states to the total proposed during each iteration.

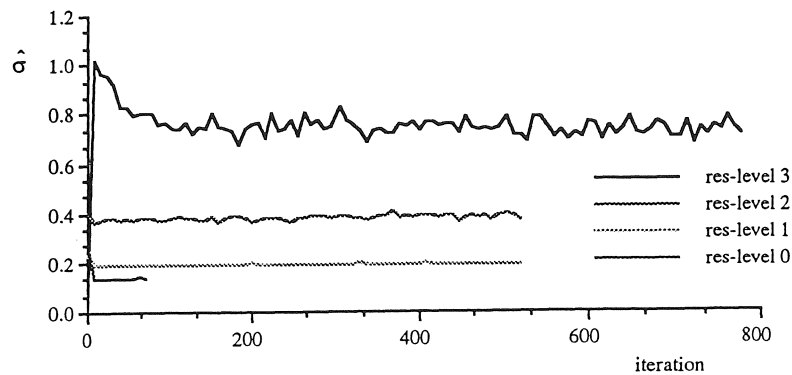


Fig. 6.8. Evolution of scale parameter estimate.

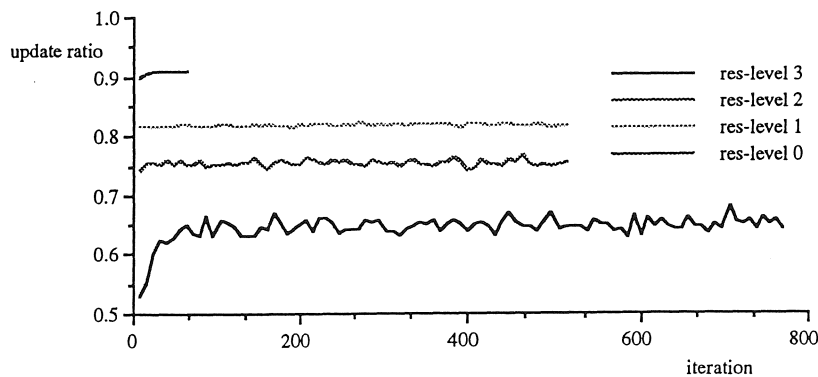


Fig. 6.9. Evolution of update ratio.

To examine the effectiveness of the multi-resolution procedure, the reconstruction at level 1 was carried out with the reference surface, rather than the reconstruction at level 2, as a starting point. The resulting surface, shown in Fig. 6.10, is significantly inferior to the multi-resolution result shown in Fig. 6.6. The evolution of the data term is shown in Fig. 6.11. Note that the energy decreases much more quickly for the multi-resolution case, because the starting point is closer to a low-energy state. (A low-energy state is a “representative” state, as it represents a “typical” sample under the posterior.) It is interesting however that during the later iterations both energies are small, even though the fixed-resolution result is clearly visually poor.

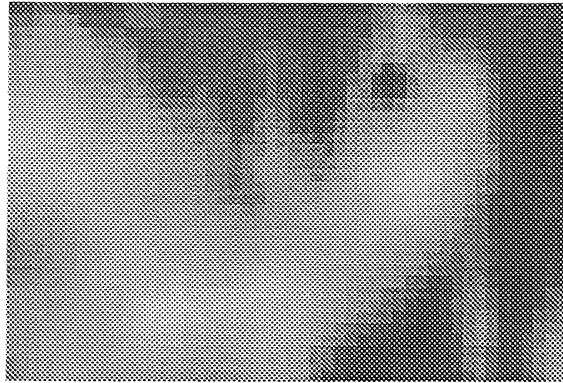


Fig. 6.10. Fixed-resolution reconstruction.

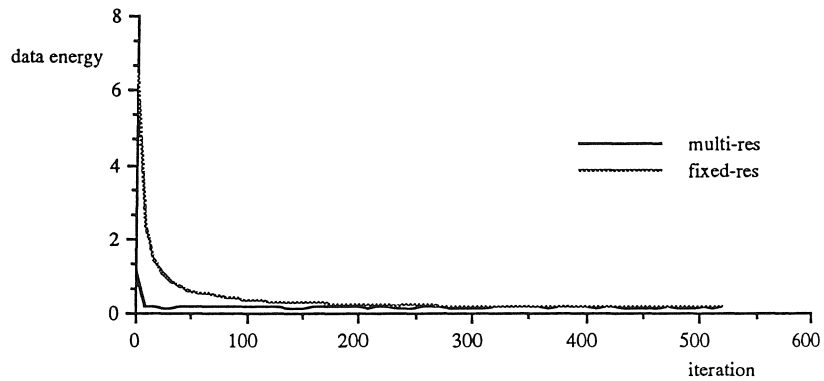


Fig. 6.11. Evolution of data energy.

Note that Fig. 6.5 is a sample from the data distribution. The model can be verified by confirming that the sample has the same visual characteristics as the Magellan data shown later. To allow comparison with the large Magellan images shown later, Fig. 6.12 shows a larger sample displayed in the same scale.

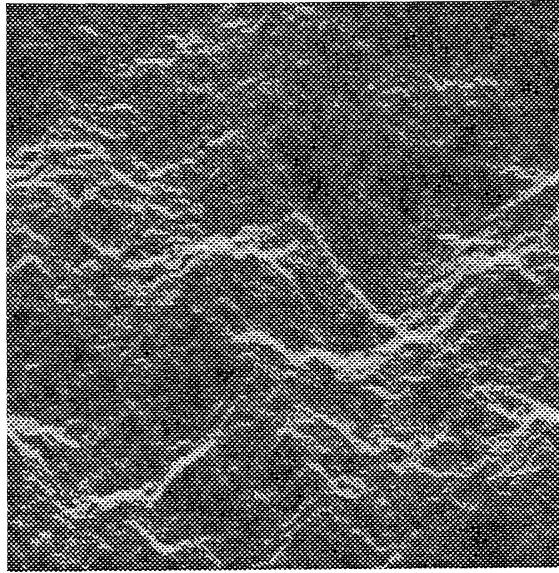


Fig. 6.12. Sample from data distribution.

Magellan Data, Freyja Montes (77°N,337°E). A 512×512 overview image in the area of Venus which is the first region chosen for experimentation is shown in Fig. 6.13. The image represents energy (in decibels) returned from the surface as a function of range and azimuth, where range increases in the downward direction on the page.

As presented in detail earlier, the image formation mechanism for radar is different than for visible-band sensing, so a radar image cannot be interpreted as an “ordinary” visible-light image, to which the human visual system is adapted. Indeed, the main benefit of radarclinometry is the production of data products which are suitable for geologic interpretation. A basic problem is the phenomenon which gives rise to layover, described earlier in Chapter 2. In the words of Wildey [4]:

The geologic interpretation of surficial expression of terrain through radar has been widely inhibited by the radar-peculiar distortion that foreshortens the extent of terrain sloped toward the radar and elongates terrain sloped away from the radar...

The distortion can be seen in Fig. 6.13, where bright snake-like patterns correspond to the faces of hills. Ground range is different than range: the entire face of a hill oriented perpendicularly to the sensor is at the same range from the sensor. The distortion can be thought of as a coordinate transformation which depends on the surface shape.

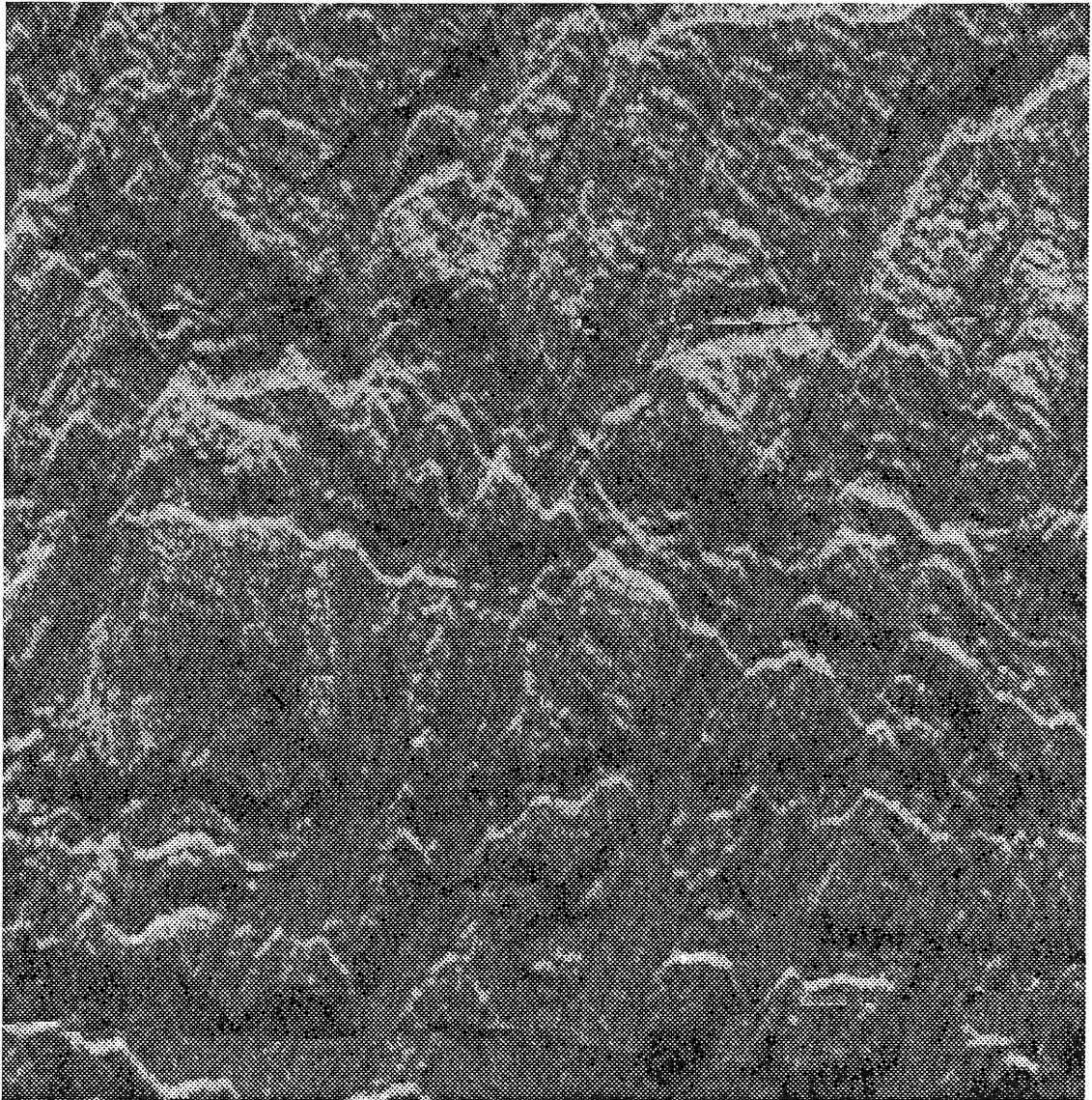


Fig. 6.13. Magellan radar image.

Magellan also collects altimetry data through microwave soundings. However, the corresponding elevation data are lower in resolution by roughly a factor of 60. As a result,

as is evident in Fig. 6.14, which shows the elevation data corresponding to Fig. 6.13, the data convey only large-scale surface properties and none of the information about small-scale surface structure which is apparent in the radar image data. It is the role of radarclinometry to reconstruct the small-scale surface features. One perspective is that the altimetry data and radar data both convey elevation information, but in different portions of the spatial frequency spectrum (see Frankot and Chellappa [1] for a discussion).



Fig. 6.14. Magellan elevation data.

Figure 6.15 shows a 96×64 portion of Freyja Montes to which the radarclinometry technique was applied. The flight and reconstruction parameters are shown in Tables 6.1 and 6.2. The image-formation scale parameter γ was estimated by $\hat{\gamma} = \sum_{s \in S} u_s / \phi_s(0)$, which is the maximum likelihood estimate in the situation that the surface function is given by the reference surface.

The reconstructed surface is shown in Fig. 6.16. The noise-free radar image of the reconstruction is shown in Fig. 6.17, which compares favorably to Fig. 6.15, as expected. Figure 6.18 is the height image based on the reconstruction, and Fig. 19 a mesh plot of the height image. Fig. 6.20 is a visible image, suitable for human interpretation, synthetically rendered from the height image, where the light source is to the top-right of the page. The scale parameter and update-ratio evolutions are shown in Figs. 6.21 and 6.22.

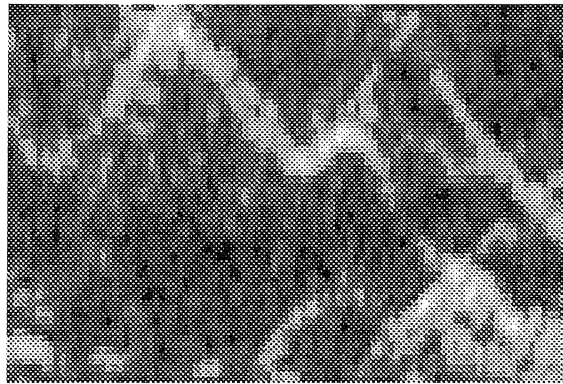


Fig. 6.15. Radar image.

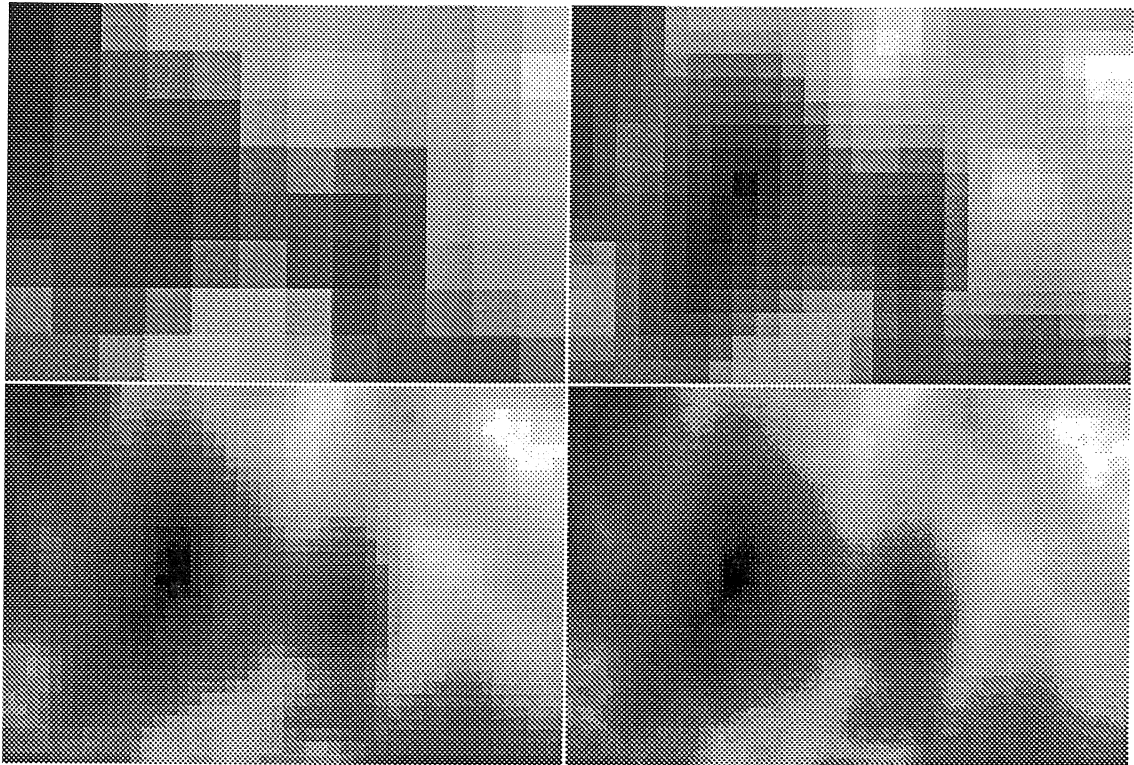


Fig. 6.16. Reconstructions; top-left: level 3; top-right: level 2; bottom-left: level 1; bottom-right: level 0.

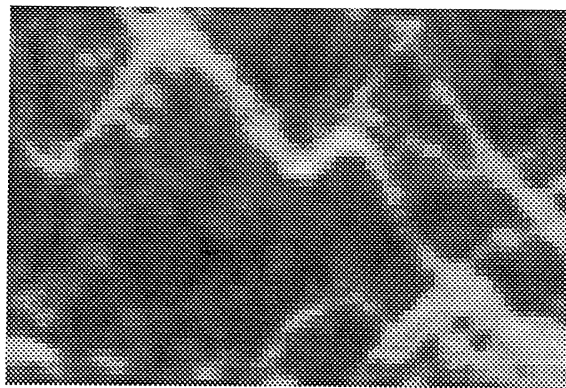


Fig. 6.17. Noise-free image of reconstruction.

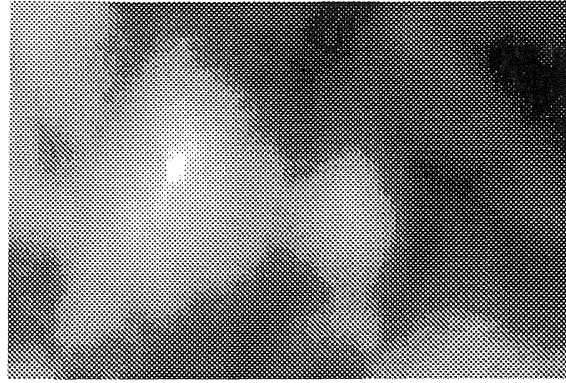


Fig. 6.18. Height image.

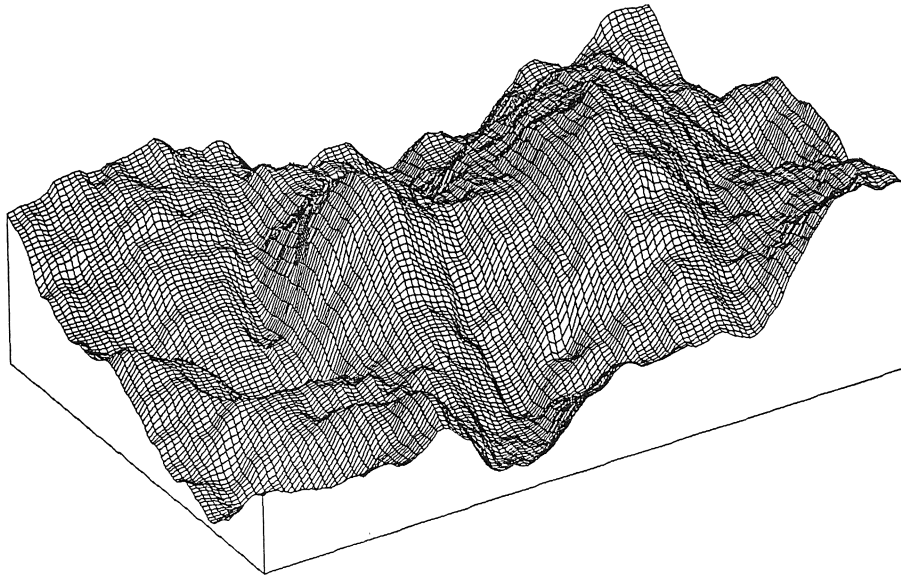


Fig. 6.19. Mesh plot of height image.

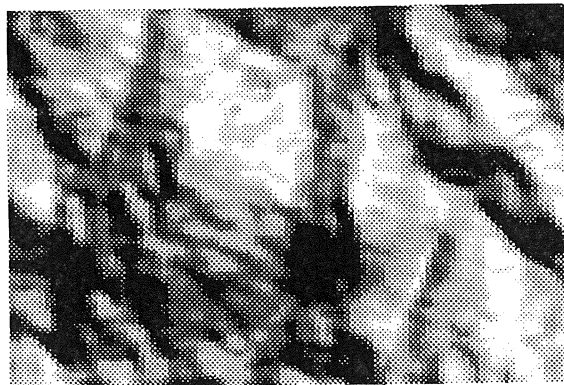


Fig. 6.20. Synthetic visual image.

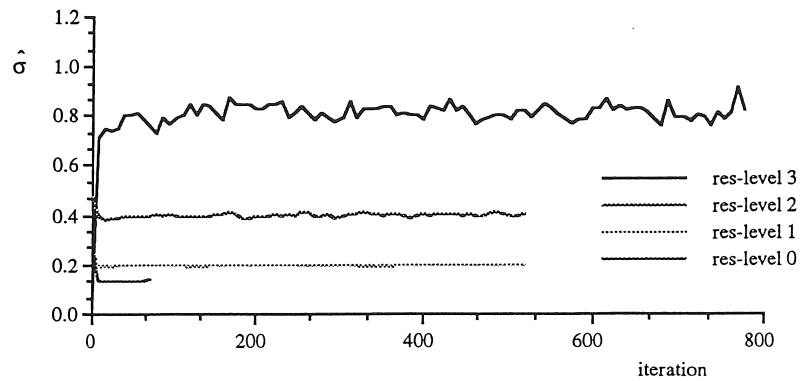


Fig. 6.21. Evolution of scale parameter estimate.

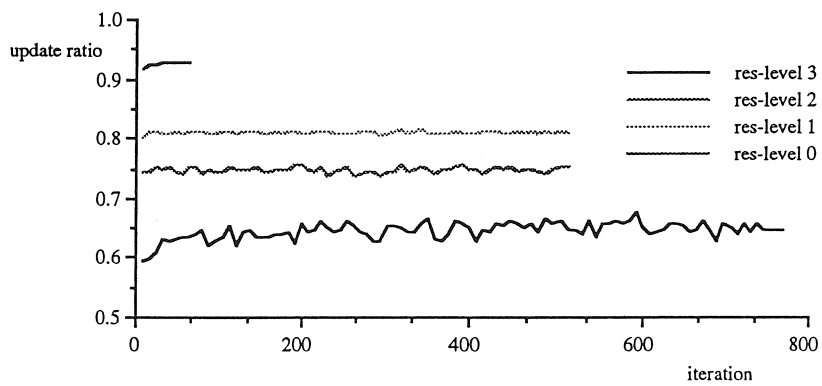


Fig. 6.22. Evolution of update ratio.

The result of processing a large data set (256×192) from Freyja Montes using the same parameters is shown in the following, where processing at the highest level of the resolution hierarchy was omitted, because of computational considerations. Figure 6.23 is the radar image, Fig. 6.24 the corresponding height image, and Fig. 6.25 the synthetic visual image.

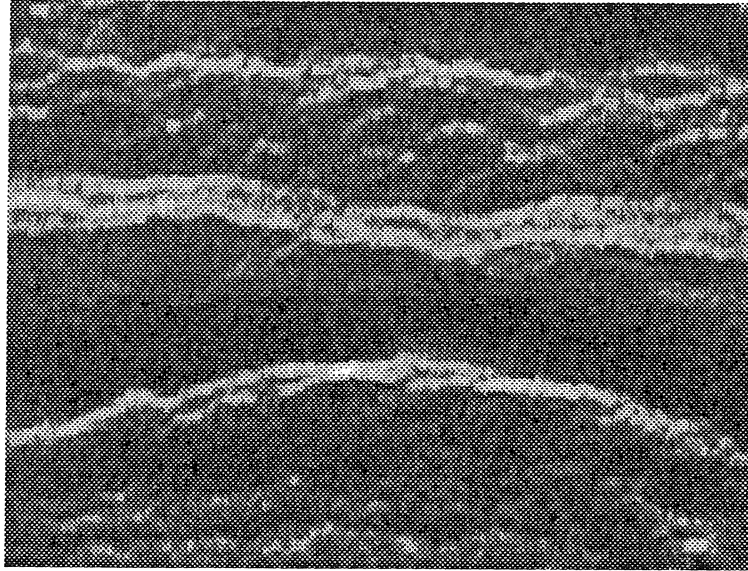


Fig. 6.23. Radar image.

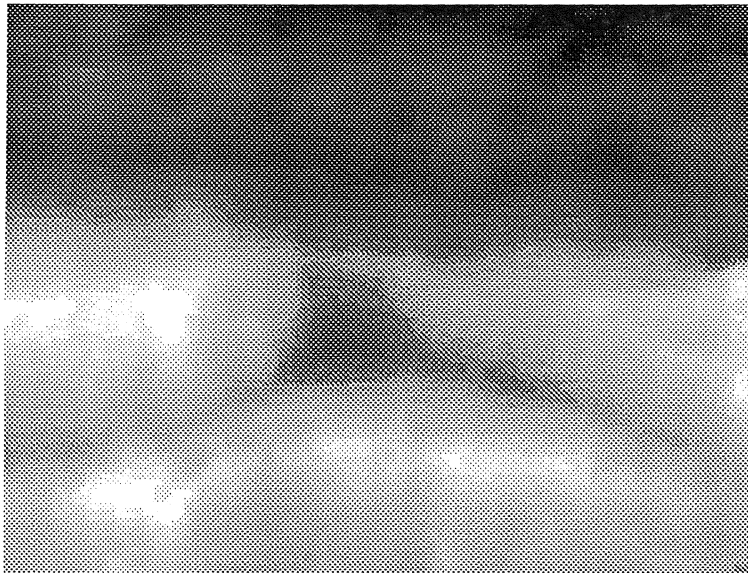


Fig. 6.24. Height image.

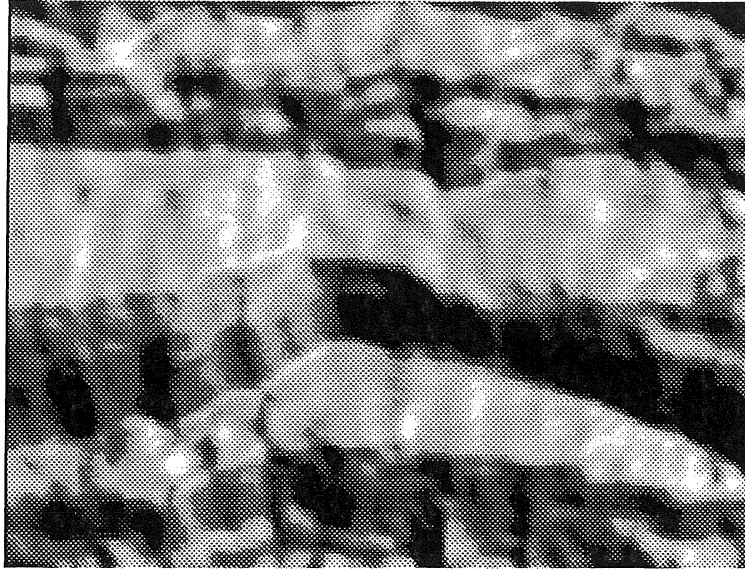


Fig. 6.25. Synthetic visual image.

Magellan Data, Alpha Regio (24°S,2°E). A (512 × 384) overview image of the second region of Venus chosen for experiments is shown in Fig. 6.26 along with a synthetic visual image, based on a reconstruction taken up to level 2, in Fig. 6.27. Figure 6.28 represents a small 96 × 64 portion to which the complete radarclinometry technique was applied. The flight parameters are shown in Table 6.3, the reconstruction parameters are as before, in Table 6.2. The reconstructed surface is shown in Fig. 6.29. The noise-free radar image of the reconstruction is shown in Fig. 6.30. Figure 6.31 is the height image based on the reconstruction, and Fig. 6.32 is the visible image.

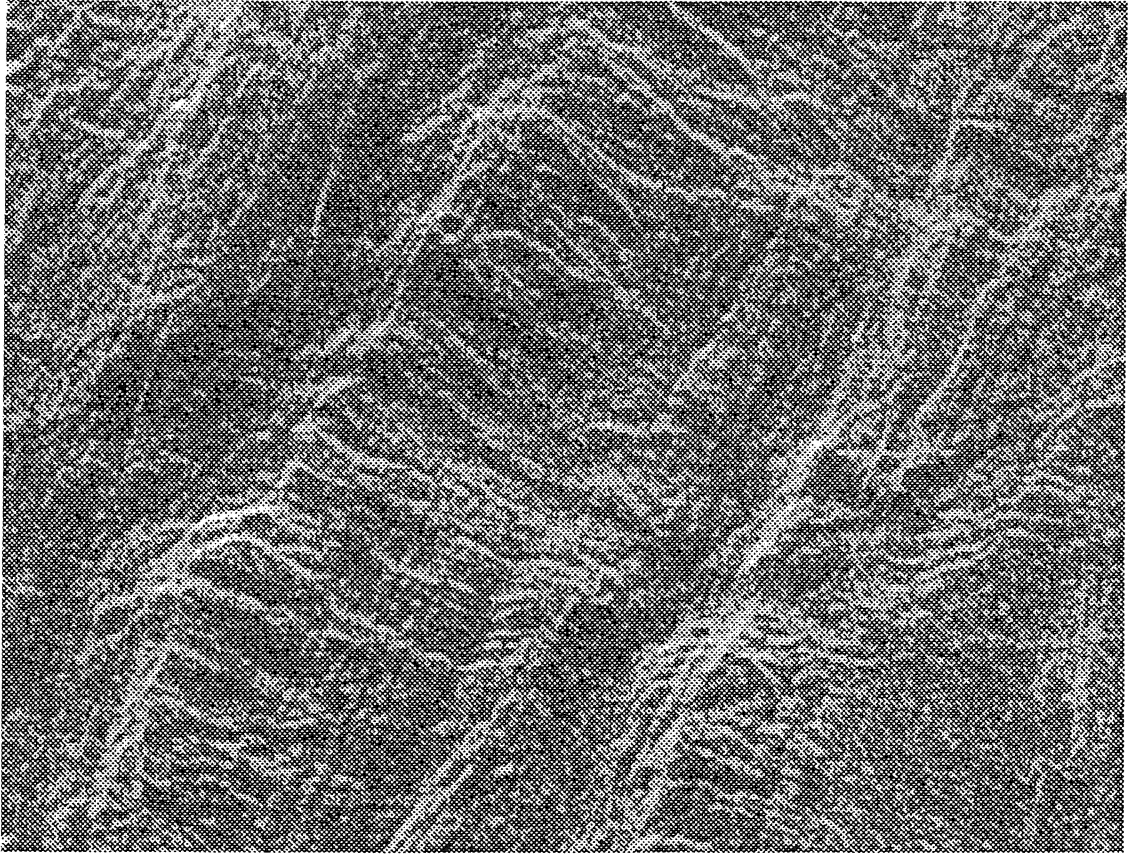


Fig. 6.26. Magellan radar image.

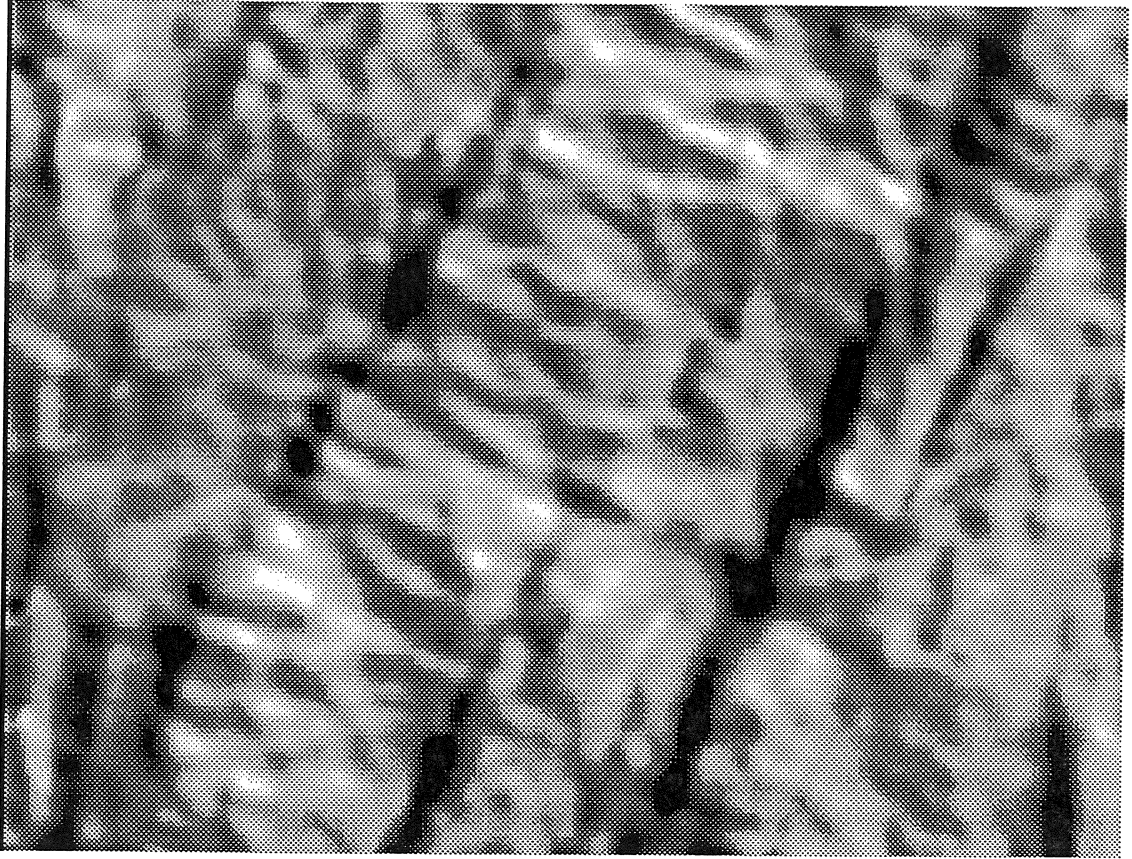


Fig. 6.27. Synthetic visual image.

Table 6.3. Flight parameters.

incidence angle (θ_0)	36°
least range	1300 km
greatest range	1309.6 km
range resolution	150 m
azimuth resolution	120 m

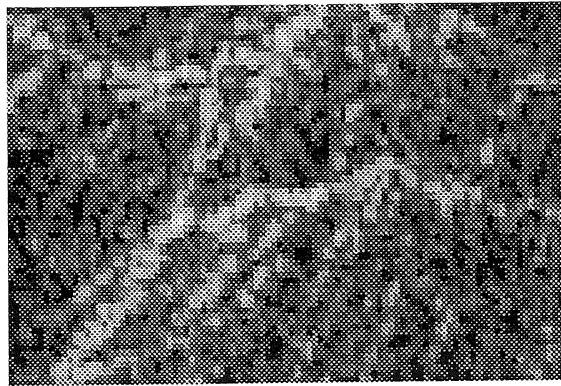


Fig. 6.28. Radar image.

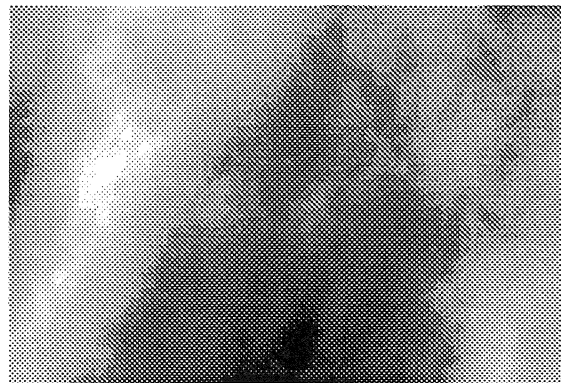


Fig. 6.29. Reconstruction.

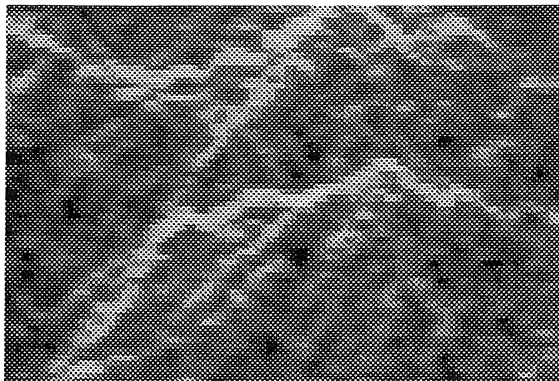


Fig. 6.30. Noise-free image of reconstruction.

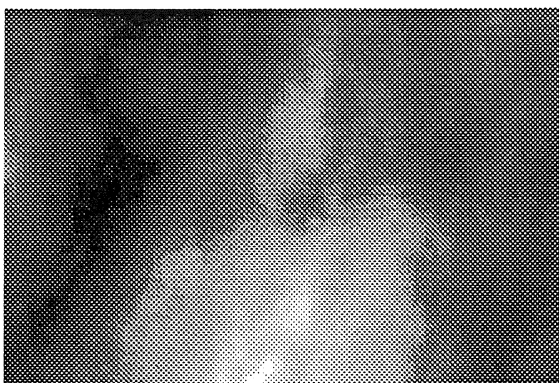


Fig. 6.31. Height image.

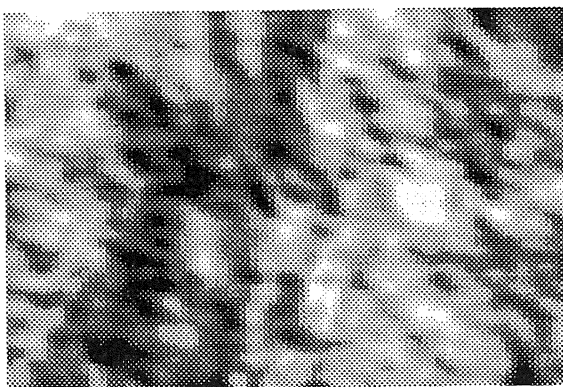


Fig. 6.32. Synthetic visual image.

Discussion

The experimental results based on uniform albedo ($\delta \equiv 1$) and Lambert reflectance are favorable. It is worth considering what is involved in designing a procedure based on the general image formation model (2.41) presented in Chapter 2. The reflectance model is not necessarily Lambert, but also the image formation operator maps not only the surface shape function ρ but also the albedo function δ to a noise-free radar image. Thus the inverse problem is more challenging, and indeed is clearly ill-posed without additional constraints. One possibility is to restrict the range of the albedo function to a small set of possibilities. In other words, the surface is assumed to be composed of only a few material types. This is a topic for further study.

References

1. R. T. Frankot and R. Chellappa, (1990). "Estimation of surface topography from SAR imagery using shape from shading techniques," *Artificial Intelligence*, 43, pp. 271-310.
2. B. Gidas, (1989). "A renormalization group approach to image processing problems," *IEEE Trans. Patt. Anal. Machine Intell.*, 11, pp. 164-180.
3. A. Rosenfeld, (1984). *Multiresolution Image Processing and Analysis*. Berlin: Springer Verlag.

BIBLIOGRAPHY

- R. W. Boyd, (1983). *Radiometry and the Detection of Optical Radiation*, New York: John Wiley and Sons.
- L. E. Franks, (1981). *Signal Theory*, Stroudsburg, Pennsylvania: Dowden and Culver.
- B. Horn and M. Brooks, (1989). *Shape from Shading*, Cambridge: MIT Press.
- D. Isaacson and R. Madsen, (1976). *Markov Chains: Theory and Applications*. New York: John Wiley and Sons.
- S. Kullback, (1959). *Information Theory and Statistics*. New York: John Wiley and Sons.
- B. B. Mandelbrot, (1983). *The Fractal Geometry of Nature*. New York: W. H. Freeman & Co.
- B. D. Ripley, (1981). *Spatial Statistics*. New York: John Wiley and Sons.
- A. Rosenfeld, (1984). *Multiresolution Image Processing and Analysis*. Berlin: Springer Verlag.
- A. N. Tikhonov and V. Y. Arsenin, (1977). *Solutions of Ill-posed Problems*. New York: John Wiley and Sons.