

Abstract

The Curve Indicator Random Field

Jonas August

2001

Can the organization of local image measurements into curves be directly related to natural image structure? By viewing curve enhancement as a statistical estimation problem, we suggest that it can. In particular, the classical Gestalt perceptual organization cues of proximity and good continuation—the basis of many current curve enhancement systems—can be statistically measured in images. As a prior for our estimation approach we introduce the curve indicator random field (CIRF). Technically, this random field is a superposition of local times of Markov processes that model the individual curves; intuitively, it is an idealized artist’s sketch, where the value of the field is the amount of ink deposited by the artist’s pen. The explicit formulation of the CIRF allows the calculation of tractable formulas for its cumulants and moment generating functional. A novel aspect of the CIRF is that contour intersections can be explicitly incorporated.

More fundamentally, the CIRF is a model of an ideal edge/line map, and therefore provides a basis for rigorously understanding real (noisy, blurry) edge/line measurements as an observation of the CIRF. This model therefore allows us to derive nonlinear filters for enhancing contour structure in noisy images. In particular, we first introduce linear, quadratic, and cubic Volterra filters for rapidly estimating the CIRF embedded in large amounts of noise. Then we derive coupled reaction-diffusion-convection integro-elliptic partial differential equations for estimating the posterior mean of the CIRF. Example computations illustrate a striking degree of noise cleaning and contour enhancement.

But the framework also suggests we seek in natural images those correlations that were exploited in deriving filters. We present the results of some edge correlation measurements that not only confirm the presence of Gestalt cues, but also suggest that curvature has a role in curve enhancement. A Markov process model for contour curvature is therefore introduced, where it is shown that its most probable realizations include the Euler spiral, a curve minimizing changes in curvature. Contour computations with curvature highlight how our filters are curvature-selective, even when curvature is not explicitly measured in the input.

The Curve Indicator Random Field

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Jonas August

Dissertation Director: Steven W. Zucker

December 2001

Copyright © 2001 by Jonas August

All rights reserved.

Contents

Acknowledgements	x
1 Introduction	1
1.1 Contributions	7
2 Markov Processes for Vision	9
2.1 Why Markov Processes?	10
2.1.1 Limitations of Markov Processes	13
2.1.2 Length Distributions	14
2.2 Planar Brownian Motion	16
2.3 Brownian Motion in Direction	17
2.4 Brownian Motion in Curvature	19
2.4.1 An Analogy	20
2.4.2 The Mathematics of the Curvature Process	21
2.4.3 What is the Mode of the Distribution of the Curvature Process?	22
2.5 Other Markov Process Models	25
2.6 A Working Formalism: Continuous-Time, Discrete-State Markov Processes	28

3	The Curve Indicator Random Field	30
3.1	Defining the CIRF	31
3.1.1	Global to Local: Contour Intersections and the CIRF	33
3.2	Moments of the Single-Curve CIRF	34
3.3	Moment Generating Functional of the Single-Curve CIRF: A Feynman-Kac Formula	39
3.3.1	Sufficient Condition For MGF Convergence: Khas'minskii's Condition	41
3.3.2	Initial and Final Weightings	41
3.4	Multiple-Curve Moment Generating Functional	44
3.4.1	The Poisson Measure Construction	44
3.4.2	Application to the MGF of the Curve Indicator Random Field for Multiple Curves	46
3.4.3	Cumulants of the CIRF	47
4	Edge Correlations in Natural Images	51
4.1	Statistical Homogeneity for Random Fields in $\mathbb{R}^2 \times \mathbb{S}$	54
4.2	Empirical Edge Correlations	55
5	Filtering Corrupted CIRFs	60
5.1	Bayes Decision Theory	62
5.1.1	Previous Estimation Problems	63
5.1.2	Loss Functions: Which Best is Best?	66
5.1.3	Minimum Mean Squared Error Filtering	68
5.1.4	Bounding the MMSE Estimate of the Binary CIRF with a Thresholded Posterior Mean of the CIRF	69
5.2	Observation Models and Likelihoods	73

5.2.1	The Gaussian Likelihood	73
5.2.2	Log-Likelihoods with Quadratic Dependence on the CIRF . .	75
5.2.3	Exploiting Empirical Distributions of Local Operators	76
5.3	The Posterior Distribution	77
5.4	The Linear Case	78
5.4.1	Why Simple Cells and Local Edge Operators are Elongated . .	80
5.5	Volterra Filters and Higher-Order Statistics	81
5.6	Deriving CIRF Volterra Filters via Perturbation Expansion	83
5.6.1	Tensors and Generalized Cumulants	85
5.6.2	Cumulants of Polynomials	89
5.6.3	The CIRF: A Tractable Special Case	91
5.6.4	Diagram Method for Evaluating the Posterior Mean	95
5.6.5	Computational Significance of Approximately Self-Avoiding Processes	101
5.7	The Biased CIRF Approximation of the Posterior	105
5.7.1	Cumulants of the Biased CIRF	106
5.7.2	Solving for the Biased CIRF	109
5.8	Comparison with Completion Fields	114
5.9	What Can One Do with a Filtered CIRF?	116
6	Numerical Methods	118
6.1	Applying the Green's Operator by Solving the Forward Equation . .	119
6.1.1	Direction Process	120
6.1.2	Curvature Process	131
6.2	Applying the Biased Green's Operator	133
6.3	Solving for the Biased CIRF Approximation of the Posterior	135

7	Results	137
7.1	Contour Filtering in $\mathbb{R}^2 \times \mathbb{S}$ with Direction	137
7.2	Contour Filtering in $\mathbb{R}^2 \times \mathbb{S} \times \mathbb{R}$ with Curvature	149
8	Conclusion	157
8.1	Future Directions	158
	Bibliography	161

List of Figures

1.1	Touching beaks or crossing curves?	2
1.2	An image of a guide wire observed fluoroscopically	4
2.1	“Lifting” a two-dimensional planar curve	11
2.2	Direction diffusions	18
2.3	Mistracking without curvature	20
2.4	Curvature diffusions for various initial curvatures	23
3.1	Similarity of contours in natural images and samples of the CIRF . .	32
3.2	Impulse response of the covariance matrix for direction process-based CIRF	49
3.3	Impulse responses of the covariance matrix for the curve indicator random field for the curvature process	50
4.1	The homogeneity of the measurement random field	54
4.2	The empirical edge covariance function of a pure noise image versus that of a natural image with contours	56
4.3	Edge covariances (continued)	57
4.4	Comparison of the correlation coefficient between collinear edges . . .	58
4.5	Edge correlations as a function of position only	59

5.1	Mis-estimation of a binary image	67
5.2	Function used in the proof of Proposition 6	72
5.3	Diagrams for representing terms of the posterior mean (5.14)	97
5.4	Diagrams with loops for representing permutations in the CIRF cumulant formula	99
6.1	Isotropy comparison	127
6.2	The discrete impulse is the sampling of a (periodic) sinc function . .	128
6.3	Comparison of solution improvement with increased spatial resolution	130
6.4	Decrease in relative residual with conjugate gradient iterations	131
6.5	Rate of improvement in solution with increase in resolution	132
7.1	Enhancing a line in the presence of noise	140
7.2	Enhancing a line in the presence of noise (continued)	141
7.3	Noise performance of direction process-based MMSE CIRF Volterra filters	142
7.4	The effect of varying the direction process parameters	143
7.5	Filtering crossing contours with direction process-based Volterra filters	145
7.6	Crossing lines are teased apart with the cubic Volterra filter	146
7.7	Fluoroscopic image of guide wire image and logical/linear negative-contrast line responses	147
7.8	Finding a guide wire	148
7.9	Finding a ship's wake	150
7.10	Filtering with the biased CIRF posterior mean approximation	151
7.11	Curvature filtering of a blood cell image	153
7.12	Curvature filtering for an angiogram	154
7.13	Filtering results on Euler spiral images	156

List of Notation

$A \times B$, cartesian product of sets A and B , 11	$R_t = R(t)$, Markov process, 10
A , likelihood matrix, 74	$T-$, left limit, 36
B , blur operator, 51, 73	T , length, 14
C , diag c , 36	$T^{(n)}$, length of n -th curve, 31
D_i , matrix with 1 at (i, i) , 106	U , CIRF, 31
$E(\gamma)$, Euler spiral functional, 25	V , CIRF (single curve), 31
$G(c) = G_c$, Green's operator with bias c , 39	$V^{(n)}$, n -th single-curve CIRF, 31
G , Green's operator, 28	W , standard Brownian motion, 16
G^* , transpose of G , 48	Z_c , biased CIRF normalizing constant, 105
H , hitting random field, 69	CGF, cumulant generating functional, 84
L , Markov process generator (discrete space), 28	\mathbb{E} , expectation, 36
M , measurement random field, 50, 61	\mathbb{E}_c , biased CIRF expectation, 105
N , noise, 73	\mathcal{F} , σ -algebra, 44
$P(t)$, transition matrix, 28	Γ , planar curve (random), 10
$P_c(du)$, biased CIRF distribution, 105	\mathcal{I} , state space (discrete), 28
$P_{(\epsilon)}(du)$, posterior CIRF distribution, 76	\mathbb{L} , Laplace transform, 38
Q , killed Markov process generator, 28	\mathcal{L} , Markov process generator (continuous space), 16
$R_t^{(n)}$, n -th Markov process, 31	\mathcal{N} , number of curves, 45

$\bar{\mathcal{N}}$, average number of curves, 45
 \mathcal{P} , polynomial operator, 88
 \mathcal{R} , continuous state-space, 120
 \mathbb{R}_+ , nonnegative real numbers, 44
 \mathbb{S} , circle, 11
 $Z_{(\epsilon)}$, posterior CIRF normalizing constant, 76
 Σ_U , CIRF covariance, 77
cum, cumulant, 47, 85
 cum_c , cumulant of biased CIRF, 107
 $\kappa^{i\cdots, j\cdots, k\cdots, \cdots}$, generalized cumulant, 85
 κ_c , cumulant of biased CIRF, 107
 \dagger , death state, 28
 $\delta(\cdot)$, Dirac δ -distribution, 16
 δ_i , discrete impulse, 106
 $\delta_{i,j}$, Kronecker δ , 36
 η , density, 48, 92
 $\text{diag } c$, diagonal matrix, 36
 ζ , Volterra filter parameter, 102
 \dot{x} , derivative $\frac{\partial x}{\partial t}$, 9
 ϵ , inverse noise variance, 102
 \tilde{x} , estimate of X , 62
 ν , final weighting, 35
 γ , planar curve, 9
 \hat{f} , Fourier transform of f , 121
 \mathfrak{i} , $\sqrt{-1}$, 9, 121
 \mathbf{I} , image, 137
 $\mathbb{1}\{\cdot\}$, indicator function, 31
 μ , initial distribution, 36
 $\langle \cdot, \cdot \rangle$, inner product, 35
 κ , curvature, 10
 λ , average length, 142
 $\mathfrak{P} \oplus \mathfrak{P}'$, edge sum, 86
 \mathfrak{P} , partition, 86
 $a \odot b$, componentwise product, 96
 ∇ , gradient, 22
 ∇_c , gradient w.r.t. c , 105
perm, permutations, 43
 rot_ϕ , rotation matrix, 53
 σ_κ , direction process standard deviation, 17, 142
 σ_{BM} , planar Brownian motion standard deviation, 16
 τ_i , average time per hit of i , 69
 θ , direction, 9
 $\mathbf{1}$, vector of ones, 40
 a , likelihood vector, 74
 c , bias vector, 35
 $d(c)$, closure measure, diagonal of Green's operator with bias c , 110
 d , Green's operator diagonal entry, 92
 e^A , matrix exponential, 28

$f * g$, convolution, 38
 g_{ij} , Green's operator entry, 28
 i, j , states (discrete), 28
 l_{ij} , Markov process generator entry, 28
 m , measurement random field realization, 61
 $p_{ij}(t)$, transition matrix entry, 28
 $r(t)$, Markov process realization, 11
 $U^{>0}$, binary CIRF, 61
 \mathbb{R} , real numbers, 9
 Σ_M , measurement covariance, 51
 Σ_N , noise covariance, 73
 Σ_U , CIRF covariance, 51
 λ , average length, 14
 α , inverse length, 14
 p_{off} , "off"-curve distribution, 75
 p_{on} , "on"-curve distribution, 75

Acknowledgements

I am indebted to my advisor, Steven Zucker, for many things. First I thank him for transforming me from a speculative explorer into (hopefully) a more directed researcher. Also, I appreciate how he has provided a place where I could pursue research relatively (in fact, almost entirely) without worry about grants, etc., and with a great deal of freedom. An unfortunately frequently-mentioned joke at computer vision conferences is “So where’s the vision?” Steve is one of the rare researchers in this field that that continues to focus on fundamental problems, even when others have abandoned their roots for more profitable or fashionable pastures. He also encourages—and supports—his students to do the same.

I would also like to thank my committee. I spent numerous pleasant afternoons having coffee with Nicolas Hengartner, who provided valuable statistical insights. James Duncan and Peter Belhumeur also helped me think through my ideas in the early stages of this research and about applications later on.

I am grateful for the numerical aid Vladimir Rokhlin provided when I confronted my filtering equations. I would especially like to thank him for suggesting the Fourier-based direct method and introducing me to the conjugate gradient method. I would like to thank Peter Caines at McGill for giving me a taste for probability. David Mumford generously agreed to be external reader for this thesis and provided careful and timely comments on an earlier draft. Kaleem Siddiqi helped me learn the ropes

of research; I can learn a lot by emulating him. Steve Haker carefully listened to and critiqued my thinking about the CIRF, spending many hours at the board with me.

The gang of grad students in computer vision at Yale made my grad school life much more enjoyable. Athinodoros Georghiades, Patrick Huggins, Ohad Ben-Shahar, Hansen Chen, Todd Zickler, Melissa Koudelka, and the round table crowd provided engagement and plenty of well-deserved teasing. I also enjoyed discussions with Fredrik Kahl and Niklas Ludke.

Sharon Taylor has been a constant source of support and affection, and she brought a lot of humanity to my strange experiences as a graduate student.

My parents Casey and Eileen August were always available to provide encouragement and even a refuge from New Haven. I also thank my brothers Justin and Jason for providing an ear or some flattery when I could use it.

Funding was provided by AFOSR.

Chapter 1

Introduction

There is a semantic gap between images and the visual structures suggested by them. On the one hand, an image is a distributed array of numbers representing, for example, a field of spatially distributed light intensity. On the other, the “objects” depicted by these images are discrete (e.g., a person, a car, or a ship’s wake) and form a kind of alphabet for our descriptions and interactions with the world (e.g., “the person on the left is approaching the car on the right,” or “the ship’s wake suggests the presence of a ship”). Bridging the gap between these qualitatively distinct sorts of descriptions—performing the signal-to-symbol transition—is at the heart of early vision. In this thesis we focus on one such transition: obtaining contour structure from images.

Most research on inferring image contours was inspired by the Gestalt psychologists [61, 52], who introduced a number of informal principles of perceptual organization to describe the perceived grouping of visual elements in images. For contours, most relevant is the principle of good continuation, which asserts that among the available ways of continuing a contour, the smoothest one should be chosen. Good continuation is used in grouping as a way of weighting associations of contour frag-

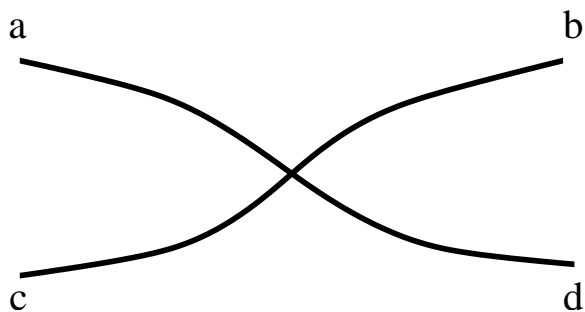


Figure 1.1: Touching beaks or crossing curves? The principle of good continuation posits that contour *a* be grouped with *d*, not *b* nor *c*. Observe in this classic demonstration that the existence of the “units” of perceptual organization, the four contours *a* through *d*, is not in question; only the association or grouping of the units must be determined. In real images, however, the contour fragments themselves are unknown and ambiguous due to noise, illumination, etc., suggesting that this grouping problem assumes that the signal-to-symbol transition is already partially solved.

ments; since these fragments must somehow already be known, grouping assumes that part of the signal-to-symbol transition has already been solved (Fig. 1.1).

The cause of this difficulty lies in the predominantly discrete nature of thinking about perceptual organization. For example, in one standard grouping demonstration, a subject is presented with a stimulus consisting of perhaps a dozen dots, and the theoretical task is to explain why the subject chose one particular partitioning of the dots among the finite set of possibilities. In illusory contour perception as well, the stimulus is a number of inducing endpoints, where one seeks to understand the (perceptual) selection of one of a finite set of arrangements. The “units” of perceptual organization here (dots, endpoints) are discrete, and typically few in number. Artificial intelligence has reinforced this view, where the units of perceptual organization become the “atoms” of LISP programs and symbol manipulation. Modern graph theoretic models of perceptual organization in computer vision have entrenched this discrete thinking, even to the extent that the formal task of perceptual organization has become identified with the grouping of a finite set of (fixed) entities or tokens. In

these discrete “symbolic” models, the underlying presence of a signal—the image—is suppressed.

Unfortunately, natural images are ambiguous, and not only in terms of the groupings of discrete units: even the existence or absence of these units is uncertain. For the inference of contour structure, this ambiguity implies that the true space of units is the uncountably infinite set of all possible curve groups; explicitly enumerating them is unthinkable. To cross this both semantic and now practical divide between images and contours, we introduce an intermediate structure—the curve indicator random field—which will allow us to bring perceptual organization closer to signal processing by formalizing what one can mean by an image of curves. In particular, we shall study and exploit the statistical structure of this field and then derive filters for extracting contour information from noisy images.

To motivate our model, consider the situation of an artist sketching a scene. Using a pen, the artist draws contours on the paper; this motion can be formalized mathematically as a curve, or mapping of a parameter, the current time in the duration of the stroke, to the point in the plane where the tip of the pen is located¹. A separate time-sequence of pen locations can be recorded for each stroke; together this set of curves provides a description of the artist’s activity. Another way of describing this situation is the drawing itself; the ink records a trace of the pen’s motion, but discards the curves’ parameterizations. The drawing might be called an “ink field,” where the value of the field at a point specifies how much ink was deposited there; this field has zero value except along the sketched contours. Combined with the randomness inherent in the act of drawing, this ink field will inspire our definition

¹Such pen motion is actually recorded in current hand-held electronic pen devices such as the Palm Pilot.

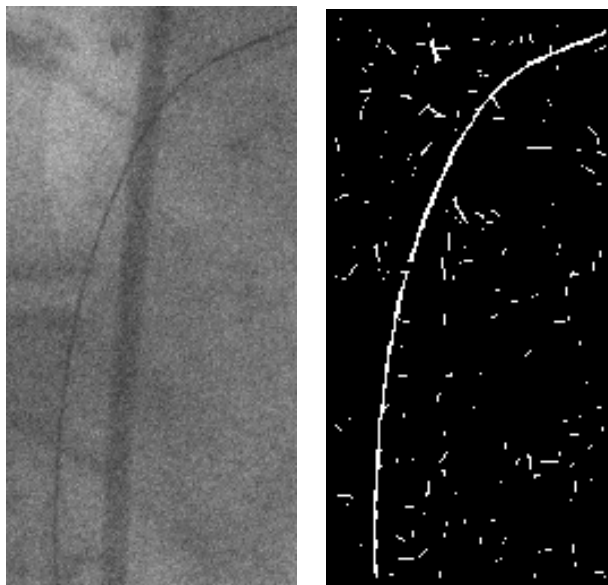


Figure 1.2: An image of a guide wire observed fluoroscopically (left) and a corresponding *non-ideal* contour image produced by thresholding a local line operator, tuned to detect the thin contour (right). Observe the numerous bogus responses away from the wire. See Fig. 7.7 for details.

of the curve indicator random field, or CIRF, to become our model of an image of contours. Estimating the CIRF given a noisy rendition of it will be our filtering task.

Given the apparently slight distinction between the set of contour points and the CIRF, it is not surprising that the CIRF has been overlooked as an object of study in vision. But there are significant advantages to making the CIRF a focus for understanding visual contours. First, it provides an exact notion of an ideal edge or line map (Fig. 1.2), satisfying the two desiderata of being (a) nonzero-valued along the true contours, and (b) zero-valued elsewhere. The CIRF therefore provides a basis for formalizing saliency [108] as the problem of estimating the CIRF. Second, unlike dictionary-based relaxation labeling [40] and line processes in Markov random fields [36, 72], the CIRF does not require an explicit and arbitrary enumeration and weighting of local field configurations to specify its joint statistical structure. Instead, by formally defining the field as a function of the curves depicted by it, the statistical properties of the field become a derivable consequence of those of the contours. Third,

being a field, the CIRF makes it possible to formalize what is even meant by observing contours under various forms of corruption. Without the CIRF, a concept of a noisy contour might mean a smooth curve made rough; with the CIRF, a noisy *image of* a contour is simply the pointwise addition of one field (the CIRF) with another (white noise), for example. Fourth, the filters we derive provide a different notion of blur scale than convolution by a Gaussian, or linear scale-space [59]. In particular, smoothing will take place along the (putative) contours. Finally, the CIRF provides a local representation of a global curve property: intersection. As ink builds up where contours cross in a sketch, so does the value of the CIRF increase at intersections.

There are two major sources of inspiration for this work. The first is the set of relaxation labeling [102] approaches to contour computation, which explicitly emphasized the need to include contextual interaction in inference, in contrast to classical edge detection techniques that regard each edge independently (followed by a linking procedure perhaps) [11]. Based on optimization [47] and game theory [78], these approaches formalize contour inference as a labeling problem using a notion of label compatibility derived from good continuation formalized using orientation [139] or co-circularity [89, 49]. The second source of inspiration is the work on curves of least energy [43] or elastica [82] and its probabilistic expression as stochastic completion fields [126]. We have introduced the CIRF to unify these works in a probabilistic framework.

This thesis is organized as follows. Because the signal-to-symbol transition is fraught with uncertainty, we shall use probabilistic techniques throughout. We begin in Chapter 2 with a review of Markov process models of the random contours that will drive the CIRF model. There we shall observe two key random models of good continuation, the first based on the local tangent direction, and the second based on curvature. Given a random contour model, we then formally define the

CIRF in Chapter 3. In the case of Markov process contour models, we then derive explicit formulas for the moments and cumulants of the CIRF, in the single-curve and multiple-curve cases, respectively, and summarize this joint statistical structure with a tractable expression for the CIRF moment generating functional. In Chapter 4 we briefly observe that our model is not arbitrary by presenting empirical edge correlations from several natural images that agree with the correlations of the CIRF.

Chapter 5 explores how we can exploit the statistics of the CIRF to derive filters by solving an estimation problem. The chapter begins with a review of Bayes decision theory, arguing for an additive loss function and leading to minimum mean-square error estimation of the CIRF given a corrupted image of it. After describing some practical observation models, or likelihoods, we then attack the core filtering problem of computing the posterior mean of the CIRF. Since we are primarily interested in filtering in high noise situations where the ambiguity is greatest, we derive linear, quadratic, and cubic Volterra filters by taking a perturbation expansion around an infinite noise level; we employ a novel diagram technique to facilitate the lengthy hand calculations. Another class of filters is derived by approximating the posterior with a more manageable distribution, the biased CIRF. As a result we obtain (discrete approximations to) a coupled pair of reaction-diffusion-convection integro-elliptic partial differential equations (PDEs) for estimating the posterior mean. Numerical techniques for computing solutions to our filters are presented in Chapter 6, and experimental results are provided in Chapter 7. Our results demonstrate a striking ability of our filters for enhancing contour structure in noisy images, including the guide wire of Fig. 1.2. The results also show that filters are orientation- and curvature-selective. We conclude and suggest future work in Chapter 8.

1.1 Contributions

In this thesis, we present a number of results and novel ideas for understanding and inferring curve structure in images. Specifically, we

- Introduce the curve indicator random field as an ideal edge/line map. We suggest that the statistical regularities of the CIRF can be exploited for enhancing noisy edge/line maps. The CIRF is applicable even if the individual contours are neither Markovian nor independent;
- Prove an accessible proof of the Feynman-Kac formula for discrete-space Markov processes, leading to the moments of the single-curve CIRF;
- Derive the cumulants and a tractable form for the moment generating functional of the multiple-curve CIRF.
- Show that the curve indicator random field is non-Gaussian;
- Introduce a curvature-based Markov process as a model of smooth contours. We prove that the most probable realization of this process is an Euler spiral;
- Formalize the problem of optimal contour inference (given a corrupted image) as the minimum-mean square estimate (MMSE) of the CIRF;
- Derive novel high-noise linear, quadratic and cubic Volterra MMSE CIRF filters with low computational complexity. The derivation exploits a novel diagram construction;
- Compute the cumulants of a biased CIRF and derived MMSE CIRF filters by approximating the CIRF posterior with a biased CIRF. This leads to new nonlinear contour filtering equations;

- Develop a rapid, accurate method for applying the Green's operator of a direction-based Markov process. Previous techniques were either very expensive or inaccurate. This technique is also extended to the curvature Markov process;
- Develop numerical techniques for solving the filtering equations;
- Apply these filters to noisy synthetic and real images, demonstrating significant contour enhancement; and
- Compute edge correlations and observe a connection to the correlations of the CIRF.

Chapter 2

Markov Processes for Vision

What class of contour models is broad enough to capture curve properties relevant for vision and yet both mathematically and computationally tractable? We start this chapter by arguing that Markov processes generally satisfy these needs, while acknowledging situations where they do not. We then review the theory of Markov processes and a process by Mumford [82] that represents the basic local geometry of a curve up to its tangent. Because curvature is also significant in vision, we then introduce a Markov process in curvature. Finally, we suggest other Markov processes that are useful for vision.

We now establish some notation and terminology. Recall that a *planar curve* is a function taking a parameter $t \in \mathbb{R}$ to a point $\gamma(t) := (x(t), y(t))$ in the plane \mathbb{R}^2 . (Also recall that the notation $A := B$ means that the symbol A is defined as the expression B , whereas $A =: B$ means the opposite.) The local *tangent* vector (\dot{x}, \dot{y}) has *direction* θ ,¹ defined as $\theta := \arg(\dot{x} + \mathfrak{i}\dot{y})$, where \mathfrak{i} is the imaginary unit $\sqrt{-1}$, the dot denotes differentiation with respect to the *arc-length* parameter t , and $\arg z$ denotes the counterclockwise angle from the real axis to z in the complex plane.

¹Directions are angles over the circle $[0, 2\pi)$, while *orientations* are angles over the half-circle $[0, \pi)$.

Without loss of generality, we assume that the tangent (\dot{x}, \dot{y}) (where it exists) has unit length, i.e., $\dot{x}^2 + \dot{y}^2 = 1$. *Curvature* κ is equal to $\dot{\theta}$, the rate of change of direction.

Into this deterministic picture we now inject the uncertainty of vision: since images are ambiguous, so too will be the structures within it. In other words, any given structure in a particular image can be viewed as a realization of a *random variable*.² In this way, any given curve γ is thus a realization of a random curve Γ .³ Since $\Gamma = \Gamma(t)$ is a one-parameter family of random planar vectors, Γ is a *random process* by definition.

2.1 Why Markov Processes?

By making randomness explicit, we clear a path for overcoming it. While the true curve is uncertain, the curve random process has a lawful *distribution* that can be exploited to support contour inferences. Put loosely, we can bias our inferences towards the probable curves and away from the rare. The rich variety of possible image curves may appear to suggest, however, that the structure of this distribution is either too simple to be useful (e.g., “uniform”) or too complex to be practical (i.e., an arbitrary distribution over the infinite-dimensional space of distributions of functions). As suggested by the Gestalt law of good continuation, the truth lies somewhere in between: although many contours can occur, most have some regular structure; e.g., object boundaries usually have a degree of smoothness. The class of Markov processes is ideal for describing such *local* statistical regularity of visual curves, while remaining tractable.

²A *random variable* may be a scalar, vector or function (i.e., a process or field). A random process has only one parameter, while a random field generally has several.

³We often will use upper and lower case letters to denote random variables and their realizations (samples), respectively. Capitals also denote operators later in this thesis.

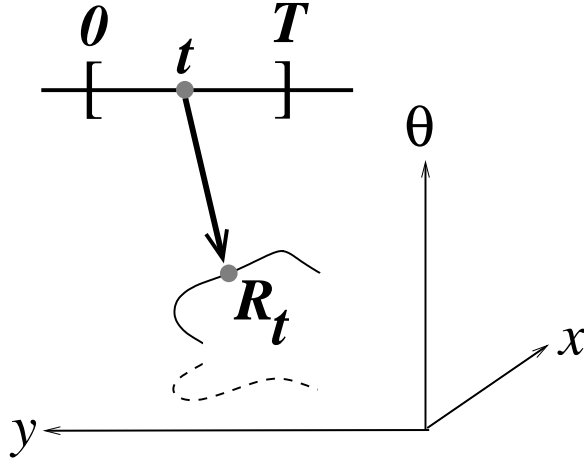


Figure 2.1: “Lifting” a two-dimensional planar curve $(X, Y)(t)$ (dotted line) to the three-dimensional space curve $R(t) = (X, Y, \Theta)(t)$ (solid line) taking values in $\mathbb{R}^2 \times \mathbb{S}$.

A random process $R = R_t = R(t)$ is *Markovian* (or a *Markov process*) if and only if the distribution of the future $\{R_t : t > t_0\}$ of the process, conditioned on the past $\{R_t : t \leq t_0\}$, is equal to the distribution of the future, conditioned only on the present (or *state*) R_{t_0} . (For details and background on Markov processes, see [86, 101].) We can construct various Markov process models R for approximating our unknown curve Γ . For example, letting the state of the process be just the planar position of the curve, we could set $R(t) = (X, Y)(t) = \Gamma(t)$ with realization $r(t) = (x, y)(t) = \gamma(t) \in \mathbb{R}^2$. But more information can be represented about the contour via the process of *lifting*, where extra components are appended to the state to take the two-dimensional curve to a curve in a higher-dimensional space (Fig. 2.1). Thus richer examples of Markov processes can be created by lifting, such as Mumford’s process with direction with the three-dimensional state $R(t) = (X, Y, \Theta)(t) = (\Gamma, \Theta)(t)$ and realization $r(t) = (x, y, \theta)(t) = (\gamma, \theta)(t) \in \mathbb{R}^2 \times \mathbb{S}$. The space $\mathbb{R}^2 \times \mathbb{S}$ is also called (x, y, θ) -space, the unit tangent bundle, and orientation space [51]. Another example of lifting, but involving a mapping from two to four dimensions, is provided by the Markov process with curvature that we introduce here

with $R(t) = (X, Y, \Theta, K)(t) = (\Gamma, \Theta, K)(t)$ with realization $r(t) = (x, y, \theta, \kappa)(t) = (\gamma, \theta, \kappa)(t) \in \mathbb{R}^2 \times \mathbb{S} \times \mathbb{R}$. These examples show that not only can the present state⁴ $R(t)$ of the Markov process represent the present value $\Gamma(t)$ of the unknown curve, it also can capture (uncertain) local differential information such as direction and curvature. In addition to the curve’s local geometry, the state can represent other local information such as blur scale [116], intensity (on one or both sides of the curve), contrast, color, texture, material information, edge classification (edge vs. line, etc.), and indeed any property that is locally definable.

Merely setting up the state vector is insufficient to characterize a Markov process: the transition probabilities are also necessary. Such probabilities describe how the present state influences the future; for visual contours, the transition probabilities therefore model how local properties extend along the curve. For example, in a world of infinitely long straight lines, the local position and direction determine those elsewhere along the curve; additionally including curvature is not necessary. For a world of very smooth contours, higher-derivative information than curvature may reliably predict nearby positions along the curves. But for fractal boundaries, even local direction would be a poor predictor of the contour nearby a point along the curve. Transition probabilities can model all of these situations, and can therefore be used to gauge the usefulness of particular local properties: those that “persist” for some distance along the curve will support the inference of these curves that we study in Chapter 5. In a sense, Markov processes both formalize the Gestalt concept of good continuation and generalize it to non-geometrical visual cues such as contrast; the transition probabilities quantitatively characterize the relative power of these cues. Markov processes are not even limited to the modeling of such generalized

⁴In a discrete-time setting there is the distinction between first-order and higher-order processes; by a suitable change of state space this distinction is irrelevant and so we consider first-order Markov processes only.

good continuations; they can also specify how good continuations break down, e.g., at discontinuities (see §2.5).

Desolneux et al take a different view on contour properties: they emphasize properties that curves do *not* have, instead of those they do [17, 18]. Instead of fixing a distribution over the smoothness of a curve (as represented via a Markov process, for example), they view smoothness as a rare (and thus meaningful) deviation from uniform randomness (contour roughness, for example).

2.1.1 Limitations of Markov Processes

While Markov processes do capture a variety of crucial properties of contours, it is important to realize that non-local cues are not explicitly included, such as closure, symmetry, simplicity (no self-intersections), and junctions (intersections between possibly distinct contours). The fundamental difficulty with all of these properties is their global nature: for a random process with any of these properties, the entire process history directly influences the future. For a Markov process, however, only the present state directly affects the future; the past has only indirect effects. Fortunately, there are several tools using which we can exploit global contour properties without abandoning Markov processes: spectral techniques, alternative shape representations, and the curve indicator random field (§3.1.1).

To study and extract closed contours, Mahamud, Williams and Thornber combined a Markov process boundary model with spectral analysis [70]. They viewed closed curves as being infinitely long; the curve repeatedly cycles through the same points. This cycling was formalized as an eigenvalue problem as follows. After extracting locally-detected edges, they constructed a reduced state-space consisting of the positions and possible directions of a subsampling of these edges. Then the eigenvector corresponding to the maximum positive eigenvalue for a (function of) the

process’s transition matrix was computed. They argued that the largest elements of this eigenvector can be interpreted as lying on a smooth closed curve, and so they iteratively removed closed curves from images by thresholding out these large elements and re-computing the maximum eigenvector for the remaining edges. In §8.1 we suggest another way to represent both symmetry and closure using a Markov process model of the skeleton representation of the boundary.

Two other global contour properties are simplicity and junctions. For example, given a parameterization of the contour, one has to traverse the entire domain of parameterization to determine whether the contour crossed itself [80]. As we shall see in §3.1.1, this global computation in terms of the parameterization becomes local in terms of the CIRF. Enforcing this constraint is discussed in §5.2, and violations of the constraint are penalized by filters that we derive in Chapter 5. Junctions represent contour crossings as well, but of distinct curves, and also are a global property of the contour parameterizations. Both of these global contour properties become local properties of a field.

So we see that while Markov processes in themselves do not convey global contour properties, they are not incompatible with global contour information when captured appropriately. We now consider one other global contour property: length.

2.1.2 Length Distributions

Not only do contours vary in their local properties, they also vary in length. While one may consider a large variety of length distributions, as a first approximation we assert a memoryless property: whether the contour continues beyond a certain point does not depend on how long it already is. We also assume that the length is independent of other aspects of the contour. This implies an exponential density $p_T(t) = \alpha \exp(-\alpha t)$ for the unknown contour length T , where $\lambda := \alpha^{-1}$ is the average

contour length. This length distribution is equivalent to a constant killing (or decay rate) α at each point in the state-space (see 2.6). This length distribution was used in [82] and [126].

One may view the exponential distribution over contour length as conservative in two respects. First, the memoryless property is an independence assumption: the probability that the curve will continue for a certain length is independent of its current length. Second, the exponential distribution can be used to represent highly ambiguous lengths. Consider what occurs as λ is increased to infinity: the slope of the density becomes increasingly flat, and so the length distribution increasingly appears uniform over any given length interval.

To model more general length distributions we can consider either (1) non-stationary Markov processes on the same state-space or (2) stationary Markov processes on an augmented state-space. As we shall require stationarity in Chapter 3, we could still take the second approach where we would add a length component to the state vector and modify the transition probabilities to monotonically increase the length component with process time. Richer length distributions would be captured by a non-constant decay rate, i.e., killing depends on the length component of the state. For example, to bias toward curves with length less than t^* one would have a small killing rate for states with length component less than t^* , and a large killing rate for states with a large length component.

With this capacity for more varied length distributions, we continue our presentation of stationary Markov processes with exponentially distributed lengths. We shall concentrate on contour geometry in continuous spaces; later we discretize in preparation for the formal development of the contour filters we derive in Chapter 5.

2.2 Planar Brownian Motion

Perhaps the simplest contour model is Brownian motion in the plane. Although we shall not be using it in our filters, Brownian motion is historically important and forms a benchmark for more complex random processes. Intuitively, it can represent the random motion of microscopic particles. Mathematically, we can use the notation of stochastic differential equations (SDEs) to describe planar Brownian motion as the random process $R(t) = (X, Y)(t)$ such that

$$dX = \sigma dW^{(1)}, \quad dY = \sigma dW^{(2)}, \quad (2.1)$$

where $W^{(1)}$ and $W^{(2)}$ are independent, standard Brownian motions⁵ and $\sigma = \sigma_{\text{BM}} > 0$ is the “standard deviation in the derivatives”. Planar Brownian motion is an example of a diffusion process and is characterized by the transition probability $p_{R(t)|R(0)}(r, t|r_0)$, which is the conditional probability density that the particle is located at r at time t given that it started at r_0 at time 0. We write the transition probability as $p = p(x, y, t) = p_{R(t)|R(0)}(r|r_0)$, where $r_0 := (x_0, y_0)$ is the initial position of the process, and they satisfy the following Fokker-Planck diffusion equation:

$$\frac{\partial p}{\partial t} = \mathcal{L}p, \quad \text{where} \quad \mathcal{L} = \frac{\sigma^2}{2} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) = \frac{\sigma^2}{2} \Delta, \quad (2.2)$$

where the initial condition is $p_{R(0)}(r) = \delta(r - r_0) = p(x, y, 0) = \delta(x - x_0)\delta(y - y_0)$, $\delta(x)$ is the Dirac δ -distribution⁶ and the spatial boundary conditions must be appropriately specified. The partial differential operator \mathcal{L} is known as the (infinitesimal) *generator* of the Markov process and fully characterizes the infinitesimal motion

⁵For reference, a standard Brownian motion W is a stationary (continuous) Markov process on \mathbb{R} where the increments are independent and increment $W_t - W_{t'}$ is a zero-mean Gaussian random variable of variance $|t - t'|$.

⁶ $\delta(x)$ is also called a Dirac δ -function.

of the particle's probability density. The other Markov processes we shall see are distinguished by their generators.

While Brownian motion has the advantage of capturing the continuity of contours (since realizations are continuous), it has two difficulties for vision. First, sample paths of Brownian motion frequently self-intersect. Not only does this not correspond to actual contours in the world, it also makes the job of filtering more difficult, as we shall see in §5.5. Second, planar Brownian motion does not represent the local geometry of contours. In particular, since Brownian paths are differentiable almost-nowhere, the tangent does not exist (except on a set of measure zero). Thus planar Brownian motion is a poor model of good continuation, elongated receptive fields in the visual cortex [45], and the observed edge statistics in natural images (see Chapter 4), where tangent information is important. The need to capture the local geometry of contours leads us to more elaborate Markov processes.

2.3 Brownian Motion in Direction

To motivate their model for local contour smoothness, first Mumford [82], and later Williams and co-workers [126], imagined a particle at $R(t) = (X_t, Y_t, \Theta_t) \in \mathbb{R}^2 \times \mathbb{S}$ whose direction Θ_t is slightly perturbed at each time instant t before taking its next step forward. Mathematically, this particle can be described by a Markov process with the stochastic differential equation

$$\frac{dX}{dt} = \sin \Theta, \quad \frac{dY}{dt} = \cos \Theta, \quad d\Theta = \sigma dW, \quad (2.3)$$

where $\sigma = \sigma_\kappa$ bounds the direction perturbations and W is standard Brownian motion (on the circle \mathbb{S}). In this way a fundamental descriptor of the local geometry

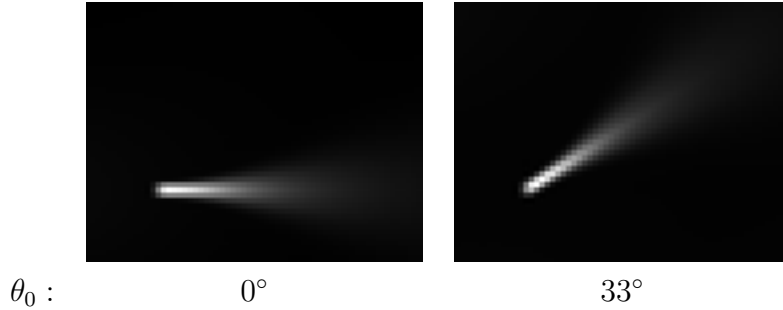


Figure 2.2: Direction diffusions. Equation (2.4) was analytically integrated over time (with exponential length weighting) and then numerically solved for a slightly blurred impulse on a $80 \times 80 \times 44$ grid, with parameters $\sigma_\kappa = 1/24$, $\lambda = 100$, and at discrete directions 0 (left) and 4 (right). Depicted is the integral over θ , cropped slightly. Unless stated otherwise, only a linear scaling is performed for all displays in this thesis; no logarithmic or other nonlinear transformation in intensity is taken. The method used responds accurately at all directions (see §6.1.1). Note that these responses are straight, and characterize how the local tangent guides nearby portions of contour. Given their initial direction, particles governed by the direction process move roughly straight ahead, in contrast to those described by the curvature process considered in §2.4 (Fig. 2.4).

of curves—the tangent $(\cos \theta, \sin \theta)$ —is captured by lifting a planar curve to $\mathbb{R}^2 \times \mathbb{S}$ and making curvature a white noise. This drives the contour randomly while ensuring that the tangent exists. This *direction process* has the transition density $p = p(x, y, \theta, t)$ (at time t) that satisfies the following Fokker-Planck diffusion equation:

$$\frac{\partial p}{\partial t} = \mathcal{L}p, \quad \text{where } \mathcal{L} = \frac{\sigma^2}{2} \frac{\partial^2}{\partial \theta^2} - \cos \theta \frac{\partial}{\partial x} - \sin \theta \frac{\partial}{\partial y}, \quad (2.4)$$

where the initial condition is $p(x, y, \theta, 0) = \delta(x - x_0)\delta(y - y_0)\delta(\theta - \theta_0)$, p is periodic in θ , and appropriate boundary conditions in x and y must be specified (see §6.1.1). Weighting by the exponential length distribution, we can illustrate these diffusions by taking an integral over time (Fig. 2.2). For a discussion of numerical techniques for the direction process, see §6.1.1. Mumford suggested the direction process as a model for occluded and illusory contours. Williams and co-workers Jacobs, Thornber, Zweck, and Wang [126, 125, 140] extensively studied this process and applied it

to modeling illusory contours using the stochastic completion field and modeling closed contours using spectral analysis (see §5.8 for comparison to the CIRF). The covariances of the curve indicator random field (Corollary 4 and Fig. 3.2) based on the direction process also resemble several formalizations of good continuation [139, 29] and horizontal interactions in the visual cortex [138].

As Mumford has shown [82], the mode of the distribution of this direction process is described by an *elastica*, or that planar curve which minimizes the following functional:

$$\int (\beta \kappa^2 + \alpha) dt, \tag{2.5}$$

where $\beta = (2\sigma^2)^{-1}$ and $\alpha = \lambda^{-1}$ is the constant killing. The elastica functional measures the bending-energy of thin splines, and was first studied by Euler [28]. Mumford derived formulas for elastica in terms of theta functions. In computer vision, elasticae were used to model illusory contours and the completion of boundary gaps [122, 43, 85].

2.4 Brownian Motion in Curvature

Inspired by Gestalt psychology [61], most previous studies of contour smoothness have focused on good continuity in orientation, that is to say, curves with varying orientation—high curvature—are rejected, and, conversely, straighter curves are enhanced. This is naturally phrased in terms of the elastica functional (2.5) on curves that minimizes curvature. Now we introduce a Markov process model that instead aims to enforce good continuation in curvature, and thus minimizes *changes* in curvature.

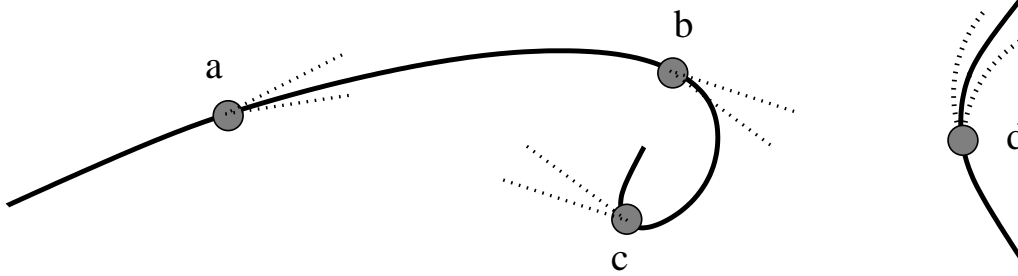


Figure 2.3: Mistracking without curvature. A bug (grey dot) attempts to track the contour, “looking” in the cone of search directions centered around its current direction. At point (a), the curve is straight and the bug is successful, although at (b), curve is veering to the right and the bug can barely still track. At (c), the curvature is so high that tracking fails. A better model would explicitly include the curvature of the contour, giving rise to a “bent” search cone (d) for the bug. The same difficulty arises in contour enhancement, which is the application considered in this thesis.

2.4.1 An Analogy

To understand why we believe that good continuation in curvature is important, imagine the situation of a bug trying to “track” the contour in Fig. 2.3. Suppose the bug is special in that it can only “search” for its next piece of contour in a cone in front of it centered around its current predicted position and direction (i.e., orientation with polarity) [133, 35]. This strategy is appropriate so long as the contour is relatively straight. However, when the bug is on a portion of the contour veering to the right, it will constantly waste time searching to the left, perhaps even mistracking completely if the curvature is too large. In estimation terms, the errors of our searching bug are correlated, a tell-tale clue that the assumption that the contour is straight is biased. A good model would only lead to an unavoidable “uncorrelated” error. We present a Markov process that models not only the contour’s direction, but also its local curvature.

It may appear that one may avoid these problems altogether by allowing a higher bound on curvature. However, this forces the bug to spend more time searching in

a larger cone. In stochastic terms, this larger cone is amounts to asserting that the current (position, direction) state has a weaker influence on the next state; in other words, the prior on contour shape is weaker (less peaked or broader). But a weaker prior will be less able to counteract a weak likelihood (high noise): it will not be robust to noise. Thus we must accept that *good continuation models based only on contour direction are forced to choose between allowing high curvature or high noise*; they cannot have it both ways.⁷

Although studying curvature is hardly new in vision, modeling it probabilistically began only recently with Zhu’s empirical model of contour curvature [135]. In [19, 138, 60] and [37, 361 ff.], measuring curvature in images was the key problem. In [58], curvature is used for smooth interpolations, following on the work on elasticae. The closest work in spirit to this is relaxation labeling [139], several applications of which include a deterministic form of curvature [89, 49]. Co-circularity (introduced in [89]) is related to curvature, and was used in Markov random field models for contour enhancement [42].

We now formally introduce our Markov process in curvature and its diffusion equation; we then present example impulse responses, which act like the “bent” search cones. We also relate the mode of the distribution for the curvature process to an energy functional on smooth curves.

2.4.2 The Mathematics of the Curvature Process

Consider the Markov process that results from making curvature a Brownian motion.

This process has state $R(t) = (X, Y, \Theta, K)(t)$, realization $r = (x, y, \theta, \kappa) \in \mathbb{R}^2 \times \mathbb{S} \times \mathbb{R}$,

⁷Observe in the road-tracking examples in [35] how all the roads have fairly low curvature. While this is realistic in flat regions such as the area of France considered, others, more mountainous perhaps, have roads that wind in the hillsides.

and can be described by the following stochastic differential equation:

$$\frac{dX}{dt} = \cos \Theta, \quad \frac{dY}{dt} = \sin \Theta, \quad \frac{d\Theta}{dt} = K, \quad dK = \sigma dW, \quad (2.6)$$

where $\sigma = \sigma_{\dot{\kappa}}$ is the “standard deviation in curvature change” (see §2.4.3) and W denotes standard Brownian motion. The corresponding Fokker-Planck diffusion equation for the bug’s probability density, is $\frac{\partial p}{\partial t} = \mathcal{L}p$, where

$$\begin{aligned} \mathcal{L} &= \frac{\sigma^2}{2} \frac{\partial^2}{\partial \kappa^2} - \cos \theta \frac{\partial}{\partial x} - \sin \theta \frac{\partial}{\partial y} - \kappa \frac{\partial}{\partial \theta} \\ &= \frac{\sigma^2}{2} \frac{\partial^2}{\partial \kappa^2} - (\cos \theta, \sin \theta, \kappa, 0) \cdot \nabla, \end{aligned} \quad (2.7)$$

where $p = p(x, y, \theta, \kappa, t) = p_{R(t)|R(0)}(r|r_0)$. Observe that this partial differential equation (PDE) describes probability transport in the $(\cos \theta, \sin \theta, \kappa, 0)$ -direction at point $r = (x, y, \theta, \kappa)$, and diffusion in κ . As in the direction process, an extra decay term [82, 126] is also included to penalize length (see §2.1.2). We have solved this parabolic equation by first analytically integrating the time variable and then discretely computing the solution to the remaining elliptic PDE. See §6.1.2 for the numerical method used to produce the example time-integrated transition densities in Fig. 2.4.

2.4.3 What is the Mode of the Distribution of the Curvature Process?

To get more insight into our random process in curvature, we now consider one of the simplest aspects of its probability distribution: its mode. Given the succinct expression of the mode of the direction process in terms of the elastica functional

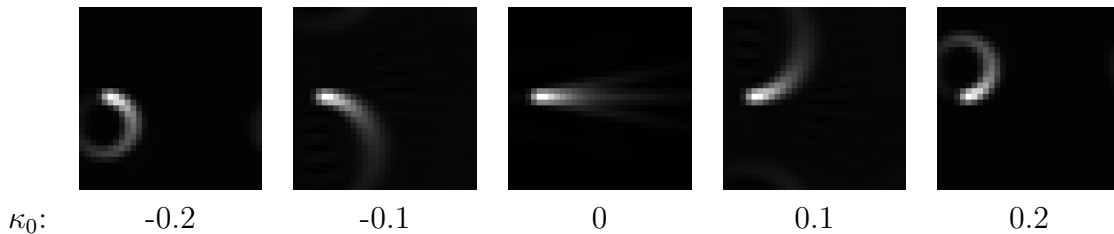


Figure 2.4: Curvature diffusions for various initial curvatures. For all cases the initial position of the particle is an impulse centered vertically on the left, directed horizontally to the right. Shown is the time integral of the transition density of the Markov process in curvature (2.7), integrated over direction θ and curvature κ ; therefore, the brightness displayed at position (x, y) indicates the expected time that the particle spent in (x, y) . Observe that the solution “veers” according to curvature, as sought in §2.4.1. Contrast this with the straight search cone in Fig. 2.2. The PDE was solved on a discrete grid of size $32 \times 32 \times 32 \times 5$, with $\sigma_{\kappa} = 0.01$ and an exponential decay of characteristic length $\lambda = 10$ (see §2.4.3 for length distribution).

(2.5), we ask: Is there a corresponding functional for the curvature process? If so, what is its form?

To answer these questions, we follow a line of analysis directly analogous to Mumford [82]. First, we discretize our random curve into N subsections. Then we write out the distribution and observe the discretization of a certain integral that will form our desired energy functional.

Recall that length T is exponentially distributed with mean $\lambda = \alpha^{-1}$. Each step of the N -link approximation to the curve therefore has length $\Delta t := T/N$. Using the definition of the t -derivatives, for example,

$$\dot{X} = \frac{dX}{dt} = \lim_{N \rightarrow \infty} \frac{X_{i+1} - X_i}{T/N},$$

we can make the approximation $X_{i+1} \approx X_i + \Delta t \dot{X}$, where i is the link index. Recalling the stochastic differential equation (2.7), we therefore let the curvature process be

approximated in discrete time by

$$X_{i+1} = X_i + \Delta t \cos \Theta_i, \quad Y_{i+1} = Y_i + \Delta t \sin \Theta_i, \quad \Theta_{i+1} = \Theta_i + \Delta t K_i,$$

where $i = 1, \dots, N$. Because Brownian motion has independent increments whose standard deviation grows with the square root $\sqrt{\Delta t}$ of the time increment Δt , the change in curvature for the discrete process becomes

$$K_{i+1} = K_i + \sqrt{\Delta t} \epsilon_i,$$

where $\{\epsilon_i\}$ is an independent and identically distributed set of 0-mean, Gaussian random variables of standard deviation $\sigma = \sigma_{\dot{\kappa}}$. Now let the discrete contour be denoted by

$$\gamma_N = \{(X_i, Y_i, \Theta_i, K_i) : i = 0, \dots, N\}.$$

Given an initial point $p_0 = (x_0, y_0, \theta_0, \kappa_0)$, the probability density for the other points is

$$p(\gamma_N | p_0) = \alpha \exp(-\alpha T) \cdot (\sqrt{2\pi}\sigma)^{-N} \exp\left(-\sum_i \frac{\epsilon_i^2}{2\sigma^2}\right),$$

which, by substitution, is proportional to

$$\exp\left[-\sum_i \frac{1}{2\sigma^2} \left(\frac{\kappa_{i+1} - \kappa_i}{\Delta t}\right)^2 \Delta t - \alpha T\right].$$

We immediately recognize $\frac{\kappa_{i+1} - \kappa_i}{\Delta t}$ as an approximation to $\frac{d\kappa}{dt} = \dot{\kappa}$, and so we conclude that

$$p(\gamma_N | p_0) \rightarrow p(\gamma | p_0) \propto e^{-E(\gamma)} \text{ as } N \rightarrow \infty,$$

where the energy $E(\gamma)$ of (continuous) curve γ is

$$E(\gamma) = \int (\beta \dot{\kappa}^2 + \alpha) dt, \quad (2.8)$$

and where $\beta = (2\sigma^2)^{-1}$ and $\alpha = \lambda^{-1}$.

Maximizers of the distribution $p(\gamma)$ for the curvature random process are planar curves that minimize the functional $E(\gamma)$, which we call the *Euler spiral functional*. When $\alpha = 0$ —when there is no penalty on length—, such curves are known as Euler spirals, and have been studied recently in [58]. A key aspect of the Euler spiral functional (2.8) is that it penalizes *changes* in curvature, preferring curves with slowly varying curvature. In contrast, the elastica functional (2.5) penalizes curvature itself, and therefore allows only relatively straight curves, to the dismay of the imaginary bug of §2.4.1. This Euler spiral energy is also loosely related to the notion of co-circularity expressed in [49]: the coefficient β would play a role of weighting the “compatibility distance” between a pair of sites (x, y, θ, κ) , curvature while the coefficient α would affect the “transport distance” between them. In other words, the integral over $\dot{\kappa}^2$ in $E(\gamma)$ is a measure of how *non-co-circular* two tangents are, while the second integral measures how far apart they are, defined as the length of γ , the smooth curve joining them.

2.5 Other Markov Process Models

At the start of this chapter we suggested the use of Markov processes for a host of contour cues other than the local geometry captured by the direction and curvature processes. The simplest way of extending a Markov process to include this extra information, say contrast C , is to add a component to the state vector and assume that

it is a Brownian motion independent of the geometry and other factors. Smoothness in C can be enforced by letting the derivative \dot{C} be a Brownian motion (or even the Ornstein-Uhlenbeck [121, 82] process suggested in 3-dimensions next), and then appending two components to the state vector (one for the contrast itself, the other for its derivative). Even dependencies among different sources of information, say intensity and geometry, can be modeled using Markov processes, but this is more complex.

While the emphasis in this thesis lies in fully exploiting the local properties of curves in the plane, the framework applies unmodified to higher dimensional images. In medical imaging of vascular structures in CAT and MRI images, one would require random process models of 3-dimensional space curves. With appropriate observation models (see §5.2), this framework could also be exploited for binocular stereo or even multiview reconstruction of space curves. In these 3-dimensional applications, one could exploit a number of properties similar to the 2-dimensional examples we have discussed. To enforce only continuity one could set the Markov process $R(t) = (X, Y, Z)(t) = (X^{(1)}, X^{(2)}, X^{(3)})(t)$ to the spatial Brownian motion

$$dX^{(i)} = \sigma dW^{(i)}, \quad i = 1, 2, 3,$$

where $\{W^{(i)} : i = 1, 2, 3\}$ are independent, standard Brownian motions. For a random 3-dimensional space curve with smoothness up to the local tangent (similar to the (planar) direction process of §2.3), Mumford suggested a Markov process obtained by integrating a Brownian motion on S^2 (the sphere of unit-length tangent vectors) [82]. A simpler model would not enforce the unit-length constraint for the tangent vector; instead, the tangent could be 3-dimensional Brownian motion. A slightly more general version of this idea includes a force (of spring constant ϵ) to

ensure that the tangent vector (the particle’s velocity $V = (V^{(1)}, V^{(2)}, V^{(3)})$) does not stray too high:

$$\frac{dX^{(i)}}{dt} = V^{(i)}, \quad dV^{(i)} = -\epsilon V^{(i)} dt + \sigma dW^{(i)}, \quad (2.9)$$

where we can eliminate the force by setting ϵ to zero. The velocity component processes $V^{(i)}$ are known as Ornstein-Uhlenbeck processes [121]. In two-dimensions, these processes were suggested for vision by Mumford [82] and related models were studied by Thornber and Williams [116]. The extension of process (2.9) to include curvature and torsion information is straightforward.

In addition to these Gaussian noise-driven Markov processes, one can construct Markov processes driven by other, heavy-tailed forms of noise. Levy processes [101] (stationary, independent-increment Markov processes) can be used as a generalized replacement for Brownian motion in all of the above Markov processes. In this way discontinuities, or “jumps” [38], in direction (i.e., corners [118]), curvature, and even contrast can be included. Zhu empirically characterized contour geometry information [135] with jumps, although in a Markov random field framework (to represent more complex relationships) instead of using an explicit Markov process. We focus on diffusion operators here both because they clearly relate to the differential geometry of curves and because they correspond to the qualitative structure of the edge correlation patterns we observe in natural images (Chapter 4). When approximated with a discrete state-space, all stationary Markov processes, including those with jumps, fit within our filtering framework.

2.6 A Working Formalism: Continuous-Time, Discrete-State Markov Processes

Having set out the kind of processes we shall ultimately use as uncertain contour models, we need to transition from continuous state-spaces to discrete ones that will support computations relevant for digital images. In this section we therefore briefly establish the notation and basic theory of discrete-state Markov processes.

With a slight abuse of notation, we let R_t , where $0 \leq t < T$, denote any stationary continuous-time Markov process taking values in a finite set \mathcal{I} of cardinality $|\mathcal{I}|$. Sites or states within \mathcal{I} will be denoted i and j . We view \mathcal{I} as a discrete approximation to our continuous state-spaces $\mathbb{R}^2 \times \mathbb{S}$, $\mathbb{R}^2 \times \mathbb{S} \times \mathbb{R}$, etc., where our contour random processes takes values. For example, (discrete) state i would represent some continuous state (x, y, θ) for the direction process, and R_t denotes the (discrete) curve Markov process. As in §2.1.2, the random variable T is exponentially-distributed with mean value λ , and represents the approximate length of the contour. To ensure finiteness of the expressions in Chapter 3, we further assume $\lambda < \infty$.

The (discrete-space) *generator* of the Markov process R_t is the $|\mathcal{I}| \times |\mathcal{I}|$ matrix $L = (l_{ij})$, and is the instantaneous rate of change of the probability transition matrix $P(t) = (p_{ij})(t)$ for R_t . Since we assume that the Markov process is stationary, the generator L does not vary with time t . One can obtain the discrete-space generator L by discretizing the continuous-space generator \mathcal{L} appropriately; details are provided in §6.1.

To enforce the exponential distribution over lifetime (length) T of each particle, we construct a *killed* Markov process with generator $Q = L - \alpha I$, where $\alpha := \lambda^{-1}$ as in §2.1.2. (Formally, we do this by adding a single “death” state \dagger to the discrete state space \mathcal{I} . When $t \geq T$, the process enters \dagger and it cannot leave.) Slightly

changing our notation again, we shall now use R_t to mean the killed Markov process with generator Q .

The Green's operator⁸ $G = (g_{ij})$ of the (killed) Markov process is the matrix $\int_0^\infty e^{Qt} dt = \int_0^\infty P(t) e^{-\alpha t} dt$, where $P(t) = e^{Lt}$ (e^A denotes the matrix exponential of matrix A). The (i, j) -entry g_{ij} in the Green's matrix represents the expected amount of time that the Markov process R_t spent in j (before death) given that the process started in i . The following is a well-known connection between the generator Q (an approximate differential operator) and the Green's operator G (an approximate integral operator):

Lemma 1. $G = -Q^{-1}$.

PROOF. Recall the standard result that $\frac{d}{dt} e^{Qt} = Q e^{Qt}$, which can be obtained by using the definition of the matrix exponential as an infinite sum, then differentiating term by term, and finally re-summing to get back a matrix exponential. We integrate this from $t = 0$ to $t = \infty$: on the left side, we get $e^{Qt} \big|_{t=0}^\infty = -I$; on the right, we get $Q \int_0^\infty e^{Qt} dt = QG$. \square

Having studied (parameterized) contour models, we remove the parameterization in the next chapter to construct our prior for contour inference from images.

⁸Formally, G is the Green's operator (or Green's matrix) when $\alpha = 0$, and is known as the resolvent, resolvent matrix, or resolvent operator for $\alpha \geq 0$.

Chapter 3

The Curve Indicator

Random Field

Given a Markov process for modeling individual contours, we now define a *curve indicator random field* (CIRF), which naturally captures the notion of an ideal edge/line map. Roughly, this random field is non-zero-valued along the true contours, and zero-valued elsewhere. The actually measured edge/line map is then viewed as an imperfect CIRF, corrupted by noise, blur, etc. Because the CIRF is not standard, our presentation will be self-contained.

Several independence assumptions play a role in our model. First, we use a memoryless property on contour length T (§2.1.2), which led to an exponential density for T . Second, we observe that independence of contours is a reasonable approximation for modeling elongated curves. Tree branch contours, for example, statistically interact primarily where they meet; due to the local process of growth, the individual branches wander largely independently. Occlusion is apparently a counter-example, although if we were interested in an “x-ray” view of contour structure, independence may still be a useful approximation. (We shall weaken the independence assumption

in §3.1.1, where contour interactions are described.) Finally, the unknown number of contours is assumed to be Poisson-distributed: in the related context of random point fields, this distribution is natural where disjoint regions contain independent numbers of points. That such assumptions are reasonable can be seen in Fig. 3.1.

3.1 Defining the CIRF

Following §2.6, we focus on continuous-time, discrete-state stationary killed Markov processes R_t , where $0 \leq t < T$. Let $\mathbb{1}\{\textit{condition}\}$ denote the (indicator) function that takes on value 1 if *condition* is true, and the value 0 otherwise. With this notation, we can define the *curve indicator random field* V for a single curve as

$$V_i := \int_0^T \mathbb{1}\{R_t = i\} dt, \quad \forall i \in \mathcal{I}.$$

Observe that V_i is the (random) amount of time that the Markov process spent in state i . In particular, V_i is zero unless the Markov process passed through site i . In the context of Brownian motion or other symmetric processes, V is variously known as the occupation measure or the local time of R_t [22, 23, 30].

Generalizing to multiple curves, we pick a random number \mathcal{N} and then choose \mathcal{N} independent copies $R_{t_1}^{(1)}, \dots, R_{t_{\mathcal{N}}}^{(\mathcal{N})}$ of the Markov process R_t , with independent lengths $T_1, \dots, T_{\mathcal{N}}$, each distributed as T . To define the multiple curve CIRF, we take the superposition of the single-curve CIRFs $V^{(1)}, \dots, V^{(\mathcal{N})}$ for the \mathcal{N} curves.

Definition 1. *The curve indicator random field U is defined as*

$$U_i := \sum_{n=1}^{\mathcal{N}} V_i^{(n)} = \sum_{n=1}^{\mathcal{N}} \int_0^{T_n} \mathbb{1}\{R_{t_n}^{(n)} = i\} dt_n, \quad \forall i \in \mathcal{I}.$$

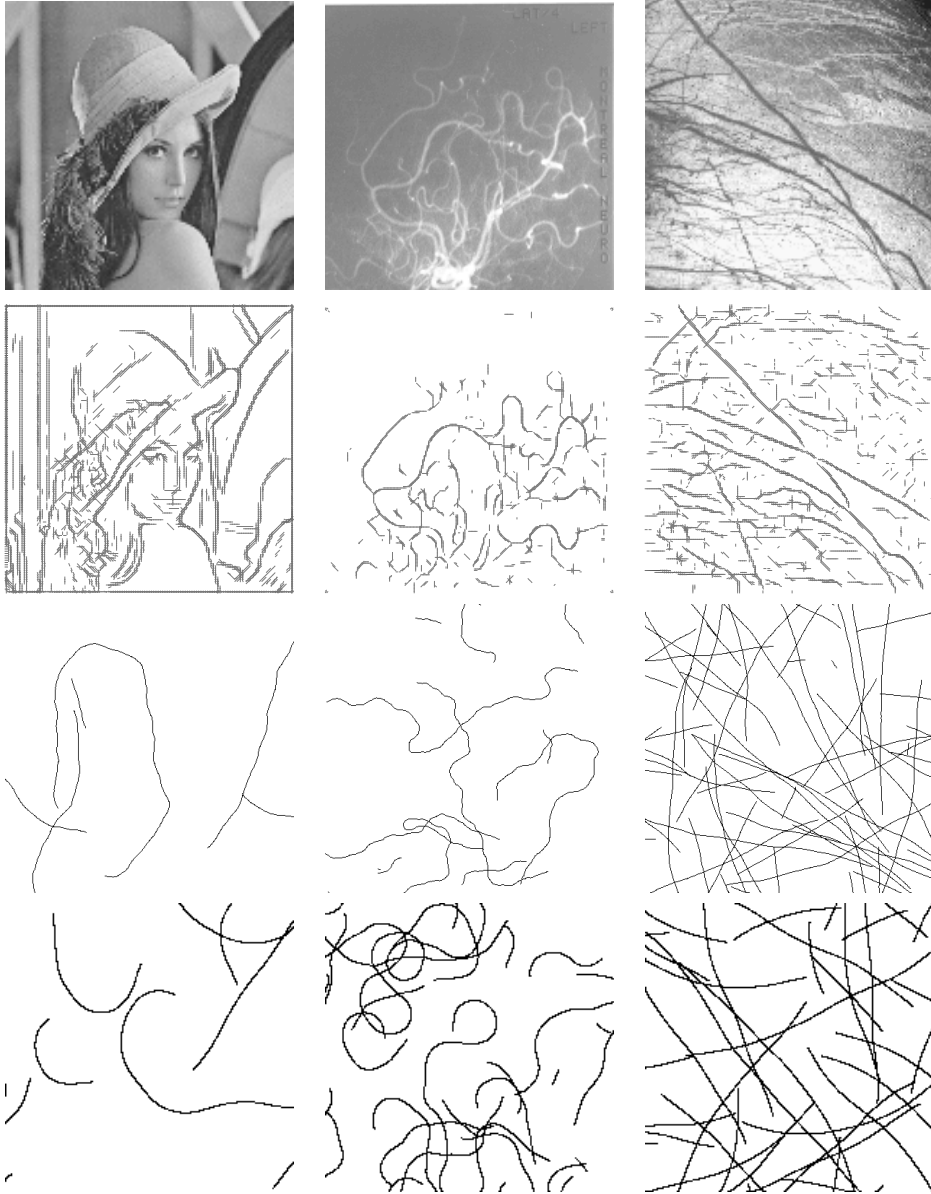


Figure 3.1: Observe the similarity of the contour structure in natural images (top row: “Lenna,” angiogram, and possible ice cracks on Jupiter moon Europa) to samples generated from the curve indicator random field (CIRF) based on the direction process (second last row: various parameter settings of direction process CIRF) and those based on the curvature process (bottom row: various parameter settings of curvature process CIRF). Although the images (top) arise from distinct physical sources, they all have contours, differing primarily in number, smoothness, and extent. The CIRF acts as a prior for contour enhancement in our framework. Notice how the CIRF resembles the qualitative structure of the local edge/line maps (second row: logical/linear edge and line operator responses [50]) corresponding to the original images (top row).

Thus U_i is the total amount of time that all of the Markov processes spent in site i . Again, observe that this definition satisfies the desiderata for an ideal edge/line map suggested in Chapter 1: (1) non-zero valued where the contours are, and (2) zero-valued elsewhere. We emphasize that the curve indicator random field is a different object than the curves used to produce it. First, the CIRF describes a random *set* of curves; each one is a Markov process. Second, and more importantly, the CIRF is a stochastic function of *space*, i.e., a random field, whereas each (parameterized) curve is a random function of time. This transformation from a set of random curves to a random field makes the CIRF an idealization of local edge/line responses, and sets the stage for contour enhancement where the probability distribution of U will become our prior for inference. See Fig. 3.1 for some samples generated by the CIRF prior.

3.1.1 Global to Local: Contour Intersections and the CIRF

In addition to representing the loss of contour parameterization, the CIRF locally captures the global contour properties of intersection and self-intersection (§2.1.1). This fact has nothing to do with the particular contour distributions used (Markovianity and independence, for example); it is a consequence of the CIRF definition. A measure of self-intersection, or the lack of simplicity of a contour, is the square of the single-curve CIRF, or

$$V_i^2 = \left(\int \mathbb{1}\{R_t = i\} dt \right) \left(\int \mathbb{1}\{R_{t'} = i\} dt' \right) = \int \left(\int \mathbb{1}\{R_{t'} = R_t = i\} dt' \right) dt.$$

To understand why V_i^2 measures self-intersection, consider the expression in parentheses on the right: *if* the process R was in site i at time t , then this expression counts the total amount of time spent in site $i = R_t$; otherwise it is zero. There are

two components to this time: (1) the (unavoidable) holding time of R_t in site i (i.e., how much time spent in i for the visit including time t) and (2) the remainder, which evidently must be the time spent in site i before or after the visit that included time t , i.e., the self-intersection time of $R_t = i$. Taking the integral over t , we therefore see that V_i^2 is a measure of self-intersection for site i over the entire contour. In continuous space and for Brownian motion, Dynkin has studied such self-intersection times by considering powers of local times for quantum field theory (which is essentially the single-curve CIRF where the indicator is replaced with a Dirac δ -distribution) [24].

A similar calculation can be carried out for the multiple-curve case U_i^2 , where we will get terms with integrals over the same curve as above (to measure self-intersection), or over a pair of distinct curves where instead we get an analogous intersection or crossing measure. The key is to recognize that a global contour property that requires checking the entire parameterization—intersection—becomes a local property (squaring) of the CIRF. If desired, we could exploit this property to enforce non-intersecting curves by penalizing non-zero values of the square of the CIRF. This would amount to changing our prior to one without our assumed independence among curves, for example. Instead of taking that direction, we first explore what statistical structure can be tractably captured without including intersections. Somewhat surprisingly however, we shall see later (§5.2) that the square of the CIRF will re-emerge in the likelihood (noise and observation model), penalizing contour intersections automatically.

3.2 Moments of the Single-Curve CIRF

Probabilistic models in vision and pattern recognition have been specified in a number of ways. For example, Markov random field models [36] are specified via clique

potentials and Gaussian models are specified via means and covariances. Here, instead of providing the distribution of the curve indicator random field itself, we derive its moment generating functional, from which all moments are then computed.

Although we are interested in the statistics of the general curve indicator random field U , we first consider the simpler, single-curve case, which we studied earlier in discrete-time [4]. The first step (Prop. 1) is to derive all the moments of the single-curve CIRF V . Then we shall summarize this result as a moment generating functional (Prop. 2).

To prepare for studying these moments, we first recognize the rather pathological nature of the CIRF: realizations of this field are zero except along the curves. While we do not develop this connection here, we observe that when the state space is a continuum, realizations of the analogous¹ CIRF would generally not even be continuous; hence we would be faced with generalized functions, or distributions (in the sense of Schwartz, not probability), where the field is probed by taking inner products with (appropriately well-behaved) “test functions” [99].

It is also convenient to use this distribution-theoretic formalism for studying the CIRF on our discrete space \mathcal{I} . To do so, we first define the inner product $\langle a, b \rangle := \sum_{i \in \mathcal{I}} a_i b_i$. We use a *bias* vector $c \in \mathbb{R}^{|\mathcal{I}|}$ to act as a “probe” or “test function” function on V by taking an inner product:

$$\langle c, V \rangle = \sum_i c_i \int \mathbb{1}\{R_t = i\} dt = \int \left(\sum_i c_i \mathbb{1}\{R_t = i\} \right) dt = \int c(R_t) dt, \quad (3.1)$$

where $c(i) = c_i$. The integral $\int c(R_t) dt$ is known as an additive functional of Markov process R_t [30]. In the following, we let $\alpha := \lambda^{-1}$ to simplify expressions. We also introduce a final weighting $\nu(R_{T-})$ on the state of the curve just before death; ν can

¹Loosely, we obtain CIRF in the continuum by replacing the indicators above with Dirac δ -distributions [3].

be used to encourage the curve to end in certain states over others.² Let $\mathbb{E}_i Z$ denote the expected value of the random variable Z given that $R_0 = i$, and let $\mathbb{E}_\mu Z$ denote the same expectation except given that $\mathbb{P}\{R_0 = i\} = \mu_i$, where μ is the distribution of the initial point of the curve. To reduce the clutter of many brackets we adopt the convention that the expectation operator applies to all multiplied (functions of) random variables to its right: e.g., $\mathbb{E}f(X)g(Y) := \mathbb{E}[f(X)g(Y)]$. We now provide a formula for the moments of $\langle c, V \rangle$, a “probed” single-curve CIRF V . The proof uses techniques that are used in statistical physics and in studying order-statistics.

Proposition 1. *The k -th moment of $\langle c, V \rangle$ with initial distribution μ and final weighting $\nu = \nu(i) = \nu_i, i \in \mathcal{I}$ is:*

$$\mathbb{E}_\mu \langle c, V \rangle^k \nu(R_{T-}) = \alpha k! \langle \mu, (GC)^k G \nu \rangle, \quad (3.2)$$

where $C = \text{diag } c$.

PROOF. We first consider the case where $\mu_j = \delta_{i,j}$, and then generalize. Recall the formula for exponentially-distributed length T : $p_T(t) = \alpha e^{-\alpha t}$. Substituting this and (3.1) into the left side of (3.2), we get:

$$\begin{aligned} \mathbb{E}_i \langle c, V \rangle^k \nu(R_{T-}) &= \mathbb{E}_i \left(\int_0^T c(R_t) dt \right)^k \nu(R_{T-}) \\ &= \alpha \mathbb{E}_i \int_0^\infty e^{-\alpha t} \left(\int_0^t c(R_{t'}) dt' \right)^k \nu(R_t) dt, \end{aligned} \quad (3.3)$$

where we have used the fact that:

$$\int f(t-) dt = \int f(t) dt,$$

²Following [30], we use the notation $T-$ to represent the left limit approaching T from below, i.e., $\nu(R_{T-}) = \lim_{t \nearrow T} \nu(R_t)$.

for piecewise continuous functions f with finite numbers of discontinuities in finite intervals. We further note that:

$$\begin{aligned} \left(\int_0^t c(R_{t'}) dt' \right)^k &= \int_0^t \cdots \int_0^t c(R_{t_1}) \cdots c(R_{t_k}) dt_1 \cdots dt_k \\ &= k! \int \cdots \int_{0 \leq t_1 \leq \cdots \leq t_k \leq t} c(R_{t_1}) \cdots c(R_{t_k}) dt_1 \cdots dt_k, \end{aligned} \quad (3.4)$$

where the second line follows because there are $k!$ orthants in the k -dimensional cube $[0, t]^k$, each having the same integral by relabeling the t_i 's appropriately. Taking integrals iteratively starting with respect to t_1 , the right side of (3.4) becomes, by induction in k :

$$k! \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} c(R_{t_1}) \cdots c(R_{t_k}) dt_1 \cdots dt_k.$$

The right side of (3.3) then becomes:

$$\begin{aligned} &\alpha k! \int_0^\infty e^{-\alpha t} \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} \sum_{i_1, \dots, i_k, j} \mathbb{P}_i \{ R_{t_1} = i_1, \dots, R_{t_k} = i_k, R_t = j \} \\ &\quad \cdot c(i_1) \cdots c(i_k) \nu(j) dt_1 \cdots dt_k \\ &= \alpha k! \sum_{i_1, \dots, i_k, j} c(i_1) \cdots c(i_k) \nu(j) \left[\int_0^\infty e^{-\alpha t} \right. \\ &\quad \cdot \left\{ \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} p_{i, i_1}(t_1) p_{i_1, i_2}(t_2 - t_1) \right. \\ &\quad \left. \left. \cdots p_{i_{k-1}, i_k}(t_k - t_{k-1}) p_{i_k, j}(t - t_k) dt_1 dt_2 \cdots dt_{k-1} \right\} dt_k \right], \end{aligned} \quad (3.5)$$

using the Markovianity and stationarity of R_t . Recalling that the formula for the convolution of two functions f and g is:

$$(f * g)(t) = \int_0^t f(\tau)g(t - \tau)d\tau,$$

we see that the expression in braces in (3.5) can be written as:

$$\begin{aligned} & \int_0^t \cdots \int_0^{t_2} p_{i,i_1}(t_1)p_{i_1,i_2}(t_2 - t_1)dt_1 \cdots p_{i_k,j}(t - t_k)dt_{k-1} \\ &= \int_0^t \cdots \int_0^{t_3} (p_{i,i_1} * p_{i_1,i_2})(t_2)p_{i_2,i_3}(t_3 - t_2)dt_2 \cdots p_{i_k,j}(t - t_k)dt_{k-1} \\ &= (p_{i,i_1} * p_{i_1,i_2} * \cdots * p_{i_{k-1},i_k} * p_{i_k,j})(t), \end{aligned}$$

which is a k -fold convolution by induction in k . Now observe that the expression in brackets in (3.5) is the Laplace transform $\mathbb{L}\{h(t)\}(\alpha)$ with respect to t , evaluated at α , of the expression $h(t)$ in braces. Therefore by using the convolution rule of the Laplace transform k times, the expression in brackets in (3.5) becomes:

$$\mathbb{L}\{p_{i,i_1}\}(\alpha) \cdot \mathbb{L}\{p_{i_1,i_2}\}(\alpha) \cdot \cdots \cdot \mathbb{L}\{p_{i_{k-1},i_k}\}(\alpha) \cdot \mathbb{L}\{p_{i_k,j}\}(\alpha).$$

But as shown earlier we know that $g_{i,j} = \int_0^\infty e^{-\alpha t} p_{i,j}(t) dt = \mathbb{L}\{p_{i,j}\}(\alpha)$, and so we can write:

$$\begin{aligned}
& \mathbb{E}_i \langle c, V \rangle^k \nu(R_{T-}) \\
&= \alpha k! \sum_{i_1, \dots, i_k, j} c(i_1) \cdots c(i_k) \nu(j) g_{i, i_1} g_{i_1, i_2} \cdots g_{i_{k-1}, i_k} g_{i_k, j} \\
&= \alpha k! \sum_{i_1} g_{i, i_1} c(i_1) \left\{ \sum_{i_2} g_{i_1, i_2} c(i_2) \cdots \left[\sum_{i_k} g_{i_{k-1}, i_k} c(i_k) \left(\sum_j g_{i_k, j} \nu(j) \right) \right] \cdots \right\} \\
&= \alpha k! \sum_{i_1} (GC)_{i, i_1} \left\{ \sum_{i_2} (GC)_{i_1, i_2} \cdots \left[\sum_{i_k} (GC)_{i_{k-1}, i_k} (G\nu)_{i_k} \right] \cdots \right\} \\
&= \alpha k! ((GC)^k G\nu)_i.
\end{aligned} \tag{3.6}$$

Since for any random variable Z we have $\mathbb{E}_\mu Z = \sum_i \mathbb{P}\{R_0 = i\} \mathbb{E}[Z | R_0 = i] = \sum_i \mu_i \mathbb{E}_i Z$, the result follows. \square

3.3 Moment Generating Functional of the Single-Curve CIRF: A Feynman-Kac Formula

To obtain a formula for the moment generating functional (MGF) of the single-curve CIRF, we first define the Green's operator $G(c) = G_c$ with (spatially-varying) bias (vector) c as the Green's operator for the killed Markov process with extra killing $-c$, i.e., having generator $Q(c) := Q + C$, where $C = \text{diag}(c_1, \dots, c_{|I|})$. The bias c behaves exactly opposite to the decay or death term α in $Q = L - \alpha I$: if c_i is positive, there is a bias towards the creation of particles at site i ; if negative, there is a bias towards killing them. Using an argument similar to the proof of Lemma 1, we know $G(c) := -Q(c)^{-1} = -(Q + C)^{-1}$. We provide an explicit condition for the invertibility

of $Q + C$ later (Prop. 3). We now compute the moment generating functional for the single-curve case using Prop. 1. This is known as the Feynman-Kac formula [22].

Proposition 2. *For all $c \in \mathbb{R}^{|I|}$ such that $|c|$ is sufficiently small,*

$$\mathbb{E}_\mu \exp\langle c, V \rangle \nu(R_{T-}) = \alpha\langle \mu, G(c)\nu \rangle.$$

PROOF. Using the power series representation of the exponential, we write:

$$\begin{aligned} \mathbb{E}_\mu \exp\langle c, V \rangle \nu(R_{T-}) &= \mathbb{E}_\mu \sum_{k=0}^{\infty} \frac{\langle c, V \rangle^k}{k!} \nu(R_{T-}) \\ &= \sum_{k=0}^{\infty} \mathbb{E}_\mu \langle c, V \rangle^k \nu(R_{T-}) / k! = \sum_{k=0}^{\infty} \alpha\langle \mu, (GC)^k G\nu \rangle \end{aligned} \quad (3.7)$$

using Prop. 1. Recalling the fact that $\sum_{k=0}^{\infty} A^k = (I - A)^{-1}$, as long as $\|A\| < 1$, the right hand side of (3.7) becomes:

$$\alpha\langle \mu, \left[\sum_{k=0}^{\infty} (GC)^k \right] G\nu \rangle = \alpha\langle \mu, (I - GC)^{-1} G\nu \rangle,$$

as long as (for some operator norm) $\|GC\| < 1$. Since $C = \text{diag}(c_1, \dots, c_{|I|})$ is diagonal, the matrix GC is simply G with the i -th column weighted by c_i . Therefore $\|GC\| < 1$ for $|c|$ sufficiently small. The result follows because $(I - GC)^{-1} = (Q^{-1}Q + Q^{-1}C)^{-1} = -(Q + C)^{-1}G^{-1}$. \square

Observe that to evaluate the Feynman-Kac formula, one must solve the linear system $(Q + C)h + \nu = 0$ for h . This equation will become a key component in the filters of Chapter 5; numerical methods are described in §6.1 and §6.2.

3.3.1 Sufficient Condition For MGF Convergence:

Khas'minskii's Condition

The proof of Prop. 2 suggests with the following condition, first studied by Khas'minskii [57], for the convergence of the Feynman-Kac formula.

Proposition 3. *The moment generating functional of the (single-curve) curve indicator random field V converges if $\|G|c|\|_\infty < 1$.*

PROOF. Using the ∞ -norm, we consider the convergence condition in the proof of Prop. 2:

$$1 > \|GC\|_\infty = \sup_i \sum_j |(GC)_{i,j}| = \sup_i (G|C|\mathbf{1})_i = \sup_i (G|c|)_i,$$

where $\mathbf{1} = (1, \dots, 1)$ and using the facts that G has positive entries and operator C is diagonal. \square

The Khas'minskii condition is easy to check: one just takes the componentwise absolute value of c (which later will be an input), and then “blur” it with the Green’s operator G . If any component of the result is greater than 1, the moment generating functional may not converge. As we shall see in §6.2, this condition will be exploited in our numerical methods for filtering.

3.3.2 Initial and Final Weightings

As the interpretation of the “final weighting” ν above may seem mysterious, we now restrict μ and ν to be finite measures satisfying the normalization constraint $1 = \langle \mu, G\nu \rangle$. (If this equality is not satisfied, one need only divide by a suitable normalizing constant.)

Corollary 1. *Suppose that the joint distribution over initial and final positions is $\mathbb{P}\{R_0 = i, R_{T-} = j\} = \mu_i g_{i,j} \nu_j$. Then the moment generating functional of V , with this joint distribution over initial and final states, is:*

$$\mathbb{E} \exp\langle c, V \rangle = \langle \mu, G(c) \nu \rangle. \quad (3.8)$$

Remark 1. *Although not studied here, it is interesting to consider the problem of finding those measures μ, ν that induce a $\mathbb{P}\{R_0, R_{T-}\}$ with prescribed marginals $\mathbb{P}\{R_0\}$ and $\mathbb{P}\{R_{T-}\}$ over initial and final states respectively.*

Since we are most interested in the case where μ and ν are spatially constant, we can forego answering this question now. Before we begin the proof of this corollary, we state a basic result.

Lemma 2. *If X and Z are random variables, and X is discrete (i.e., X can only take on one of an at most countable number of values x), then:*

$$\mathbb{E}[Z|X = x] = \frac{\mathbb{E}Z \mathbb{1}\{X = x\}}{\mathbb{P}\{X = x\}}.$$

PROOF. We compute:

$$\begin{aligned} \mathbb{E}[Z|X = x] &= \int z \mathbb{P}\{Z \in dz | X = x\} = \int z \mathbb{P}\{Z \in dz, X = x\} / \mathbb{P}\{X = x\} \\ &= \int \sum_{x'} z \mathbb{1}\{x' = x\} \mathbb{P}\{Z \in dz, X = x'\} / \mathbb{P}\{X = x\} = \mathbb{E}Z \mathbb{1}\{X = x\} / \mathbb{P}\{X = x\}. \end{aligned}$$

□

PROOF OF COROLLARY 1. Observe that $\mu_i g_{i,j} \nu_j$ is indeed a distribution. We now compute the result:

$$\begin{aligned}
\mathbb{E} \exp\langle c, V \rangle &= \sum_{i,j} \mathbb{P}\{R_0 = i, R_{T-} = j\} \mathbb{E}[\exp\langle c, V \rangle | R_0 = i, R_{T-} = j] \\
&= \sum_{i,j} \mu_i g_{i,j} \nu_j \mathbb{E}_i[\exp\langle c, V \rangle | R_{T-} = j] \\
&= \sum_{i,j} \mu_i g_{i,j} \nu_j \mathbb{E}_i[\exp\langle c, V \rangle \mathbb{1}\{R_{T-} = j\}] / \mathbb{P}_i\{R_{T-} = j\},
\end{aligned}$$

using Lemma 2. But using Prop. 2 we see that:

$$\mathbb{P}_i\{R_{T-} = j\} = \mathbb{E}_i \exp(0, V) \mathbb{1}\{R_{T-} = j\} = \alpha g_{i,j},$$

and therefore:

$$\begin{aligned}
\mathbb{E} \exp\langle c, V \rangle &= \alpha^{-1} \sum_{i,j} \mu_i \nu_j \mathbb{E}_i[\exp\langle c, V \rangle \mathbb{1}\{R_{T-} = j\}] \\
&= \alpha^{-1} \mathbb{E}_\mu \left[\exp\langle c, V \rangle \left(\sum_j \nu_j \mathbb{1}\{R_{T-} = j\} \right) \right] \\
&= \alpha^{-1} \mathbb{E}_\mu [\exp\langle c, V \rangle \nu(R_{T-})].
\end{aligned}$$

The result follows after another application of Prop. 2 to the above expectation. \square

The next corollary shows all of the (joint) moments of V . Let perm_k denote the set of permutations of the integers $1, \dots, k$.

Corollary 2. *If $k \geq 1$, the k -th (joint) moment of V at sites i_1, \dots, i_k is:*

$$\mathbb{E} V_{i_1} \cdots V_{i_k} = \sum_{i,j} \mu_i \nu_j \sum_{a \in \text{perm}_k} g_{ii_{a_1}} g_{i_{a_1} i_{a_2}} \cdots g_{i_{a_{k-1}} i_{a_k}} g_{i_{a_k} j}. \quad (3.9)$$

PROOF. Take partial derivatives of (3.8) with respect to c_{i_1}, \dots, c_{i_k} and evaluate them at $c = 0$. The only nonzero terms come from differentiating an expression proportional to (3.6). \square

In our earlier work on the curve indicator random field for *discrete-time* Markov processes [4], we arrived at a similar result, except under the condition that the sites i_1, \dots, i_k are *distinct*. In observing the connection to the Feynman-Kac formula, we have overcome this limitation and can now summarize the single-curve CIRF with its moment generating functional. The weighting over final states (the other end of the contour) is also new.

3.4 Multiple-Curve Moment Generating Functional

In order to model more than one curve in an image, we need a joint distribution over both the number of curves and the curves (and the corresponding CIRFs) themselves. To make our computations concrete, we adopt a Poisson distribution over the number \mathcal{N} of curves, and assume conditional independence of the curves given \mathcal{N} . To compute the moment generating functional of this (multiple-curve) CIRF as a Poisson distribution over (single-curve) CIRFs, we first consider the general case of Poisson “point” process given a distribution over each point, where point will be interpreted as an entire single-curve CIRF.

3.4.1 The Poisson Measure Construction

We begin with a finite measure³ $P : \mathcal{F} \rightarrow \mathbb{R}_+$ over the measure space (Ω, \mathcal{F}) , where \mathcal{F} is a σ -algebra, and \mathbb{R}_+ denotes the nonnegative real numbers. Intuitively, the

³A finite measure can always be normalized to a probability distribution because $P(\Omega) < \infty$. In particular, $\tilde{P}(\omega) := P(\omega)/P(\Omega)$ is a (normalized) probability distribution over ω .

finite measure P is the (un-normalized) distribution over “points” $\omega \in \Omega$, where in this thesis ω is a curve realization (i.e., a Markov process realization) and Ω is the set of all possible curves. We shall now define a probability distribution over random configurations $\boldsymbol{\omega} = (\omega_1, \dots, \omega_{\mathcal{N}}) \in \text{Con}(\Omega) := \{\Omega^0 = \emptyset, \Omega^1 = \Omega, \Omega^2 = \Omega \times \Omega, \Omega^3, \dots\}$, where each ω_n is a curve in Ω and \mathcal{N} is the random number of curves. In our context, Ω^0 is the 0-curve configuration (no curves), Ω^1 are the one-curve configurations (the set of single curves), Ω^2 are the two-curve configurations (the set of pairs of curves), and so on. We now compute the *Poisson point measure* via its expectation $\mathbb{E}F$ on any (measurable) function $F : \text{Con}(\Omega) \rightarrow \mathbb{R}$ (clearly this defines a probability distribution for we could take F as an indicator over any (measurable) subset of $\text{Con}(\Omega)$ to get its probability).

Proposition 4. *Suppose \mathcal{N} is a Poisson deviate with mean $P(\Omega)$. Further suppose that the points $\omega_1, \dots, \omega_n$ are (conditionally) independent and identically distributed with $P(\cdot)/P(\Omega)$, given $\mathcal{N} = n$. Then:*

$$\mathbb{E}F := \sum_{n=0}^{\infty} \frac{e^{-P(\Omega)}}{n!} \int_{\Omega^n} F(\omega_1, \dots, \omega_n) P(d\omega_1) \cdots P(d\omega_n). \quad (3.10)$$

PROOF. We need only take a conditional expectation and recall the formula for the Poisson distribution, as follows:

$$\begin{aligned} \mathbb{E}F &= \mathbb{E}(\mathbb{E}[F|\mathcal{N}]) \\ &= \mathbb{E}(\mathbb{E}F(\omega_1, \dots, \omega_{\mathcal{N}})) \\ &= \mathbb{E} \left(\int_{\Omega^{\mathcal{N}}} F(\omega_1, \dots, \omega_{\mathcal{N}}) (P(d\omega_1)/P(\Omega)) \cdots (P(d\omega_{\mathcal{N}})/P(\Omega)) \right) \\ &= \sum_{n=0}^{\infty} \frac{e^{-P(\Omega)} P(\Omega)^n}{n!} \left(P(\Omega)^{-n} \int_{\Omega^n} F(\omega_1, \dots, \omega_n) P(d\omega_1) \cdots P(d\omega_n) \right). \end{aligned}$$

The result follows. □

The above presentation of the Poisson point measure is based on Dynkin [22].

3.4.2 Application to the MGF of the Curve Indicator

Random Field for Multiple Curves

We now finally consider the joint distribution over many curves. Suppose there are $\bar{\mathcal{N}}$ contours on average, and that μ and ν are finite measures on \mathcal{I} (vectors in $\mathbb{R}^{|\mathcal{I}|}$), characterizing the initial and final positions, respectively, of the Markov processes $\{R_t^{(n)} : n = 1, \dots, \mathcal{N}\}$. As before, these measures satisfy the normalization constraint $\langle \mu, G\nu \rangle = 1$. For general-purpose contour enhancement, we typically have no a-priori preference for the start and end locations of each contour, and so we would set these measures proportional to the constant vector $\mathbf{1} = (1, \dots, 1)$. One can show that by letting $\mu_i = |\mathcal{I}|^{-1}$, $\nu_i = \lambda^{-1}$, $\forall i \in \mathcal{I}$, the above constraint is satisfied. We now state and prove the key theoretical result of this thesis, which is most closely related to the work by Dynkin [22].

Proposition 5. *Suppose the initial distribution μ and final weighting ν satisfy the normalization constraint $\langle \mu, G\nu \rangle = 1$. Then the moment generating functional of the curve indicator random field U is*

$$\mathbb{E} \exp \langle c, U \rangle = \exp \langle \mu, \bar{\mathcal{N}}(G(c) - G)\nu \rangle.$$

PROOF. To take advantage of the Poisson point measure construction, we let ω be a realization of the killed Markov process $R_t, t \in [0, T-)$, such that the finite measure $P = P(\omega)$ is the probability distribution for ω but multiplied by the constant $\bar{\mathcal{N}}$, i.e., $P(\Omega) = \bar{\mathcal{N}}$. Let $F := \exp \langle c, U \rangle = \exp \sum_{n=0}^{\mathcal{N}} \langle c, V^{(n)} \rangle = \prod_{n=0}^{\mathcal{N}} \exp \langle c, V^{(n)} \rangle$, where

$V^{(n)}$ is a function of ω_n . Applying (3.10) we obtain:

$$\begin{aligned}
\mathbb{E} \exp\langle c, U \rangle &= \sum_{n'=0}^{\infty} \frac{e^{-P(\Omega)}}{n'!} \int_{\Omega^{n'}} \prod_{n=0}^{n'} \exp\langle c, V^{(n)} \rangle P(d\omega_1) \cdots P(d\omega_{n'}) \\
&= \sum_{n'=0}^{\infty} \frac{e^{-P(\Omega)}}{n'!} \prod_{n=0}^{n'} \left(\int_{\Omega} \exp\langle c, V^{(n)} \rangle P(d\omega_n) \right) \\
&= e^{-P(\Omega)} \sum_{n'=0}^{\infty} \frac{1}{n'!} \prod_{n=0}^{n'} \left(\int_{\Omega} \exp\langle c, V^{(1)} \rangle P(d\omega_1) \right),
\end{aligned}$$

since $V^{(1)}, \dots, V^{(n')}$ are identically distributed. But then the latter integral is $\bar{\mathcal{N}} \mathbb{E} \exp\langle c, V \rangle$, and so the above sum becomes:

$$\sum_{n'=0}^{\infty} (\bar{\mathcal{N}} \mathbb{E} \exp\langle c, V \rangle)^{n'} / n'! = \exp(\bar{\mathcal{N}} \mathbb{E} \exp\langle c, V \rangle).$$

So using $P(\Omega) = \bar{\mathcal{N}} \langle \mu, G(\mathbf{0}) \nu \rangle$ and Prop. 1, we conclude:

$$\mathbb{E} \exp\langle c, U \rangle = \exp(\bar{\mathcal{N}} \langle \mu, G(c) \nu \rangle - \bar{\mathcal{N}} \langle \mu, G\nu \rangle).$$

□

3.4.3 Cumulants of the CIRF

While Prop. 5 may seem abstract, it is actually very useful. First observe that how similar in form it is to the single-curve case. More importantly, with Prop. 5 we can compute the higher-order cumulants [66, 84] of U (recall that the moments define the cumulants and vice versa):

Corollary 3. *If $k \geq 1$, the k -th (joint) cumulant of the curve indicator random field U at sites i_1, \dots, i_k is*

$$\text{cum}\{U_{i_1}, \dots, U_{i_k}\} = \bar{\mathcal{N}} \sum_{i,j} \mu_i \nu_j \sum_{a \in \text{perm}_k} g_{i a_1} g_{i a_1 a_2} \cdots g_{i a_{k-1} i a_k} g_{i a_k j}. \quad (3.11)$$

PROOF. Since the cumulant generating functional of U , which is the natural logarithm of the moment generating functional of U , differs from the moment generating functional of V by an additive constant, we use (3.9) with no further work. \square

The cumulant formula has a simple interpretation. First recall that the Green's operator entry g_{ij} is the expected amount of time spent by R_t in state j given that it started in i . For any ordering of the k points we take the product of the g_{ij} 's for the successive points in order (the first and last factors deal with the initial and final points). Since the contour could have passed through the points in any order, all permutations must be considered.

Letting G^* denote the transpose of G , we can rephrase corollary 3 to show the mean and covariance of the CIRF [3].

Corollary 4. *Suppose that $\mu_i = |\mathcal{I}|^{-1}$, $\nu_i = \lambda^{-1}$, $\forall i \in \mathcal{I}$, and let $\eta = \bar{\mathcal{N}}\lambda|\mathcal{I}|^{-1}$. The mean of the curve indicator random field U is $\mathbb{E} U_i = \eta, \forall i \in \mathcal{I}$. The covariance matrix of U is $\text{cov } U = \eta(G + G^*)$.*

One ‘‘column’’ of the covariance matrix for the direction process is illustrated in Fig. 3.2, by taking its impulse response. While the covariance pattern for the direction process expresses good continuation as local straightness, the covariance pattern for the curvature process captures the local circularity of contours. See Fig. 3.3 for several ‘‘columns’’ of the covariance matrix for the curvature process.

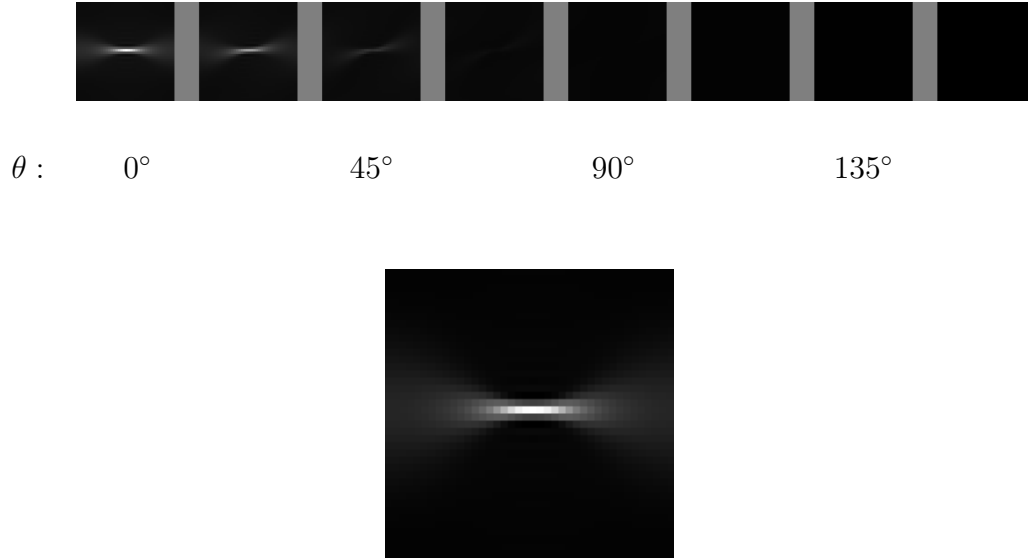


Figure 3.2: Impulse response of the covariance matrix (Corollary 4) for the curve indicator random field for the direction process in (x, y, θ) . Impulse is located at image center, with direction 0° , then Gaussian blurred with radius 1.0. Parameters are $\sigma_\kappa = 1/8, \lambda = 200, \tilde{\mathcal{N}} = 10$. (Top) Stack of images, arranged left to right, each for a different direction. (Bottom) Sum over θ of stack above, producing single image in (x, y) . Observe the elongation of the response along the horizontal direction, capturing the smoothness and length of the contours. In §5.4 we shall see that the CIRF covariance will become a linear filter for enhancing contour images. Note that the covariance impulse responses resemble compatibility fields used in relaxation labeling [139, 49] and in more recent models for horizontal connections in primary visual cortex of primates [29].

Note that by Corollary 3, the cumulants of U of order greater than two are generally *not* zero, proving the following:

Corollary 5. *The curve indicator random field U is non-Gaussian.*

Although non-Gaussianity is often difficult to handle, in our case it is easier because we have an explicit, tractable expression of the CIRF’s moment generating functional. We shall directly exploit the structure of the statistics of the CIRF in the filters we develop in Chapter 5. Before embarking on that large project, we first pause to verify the plausibility of the CIRF as a model of visual contour structure by presenting the empirical covariances of edges in natural images.

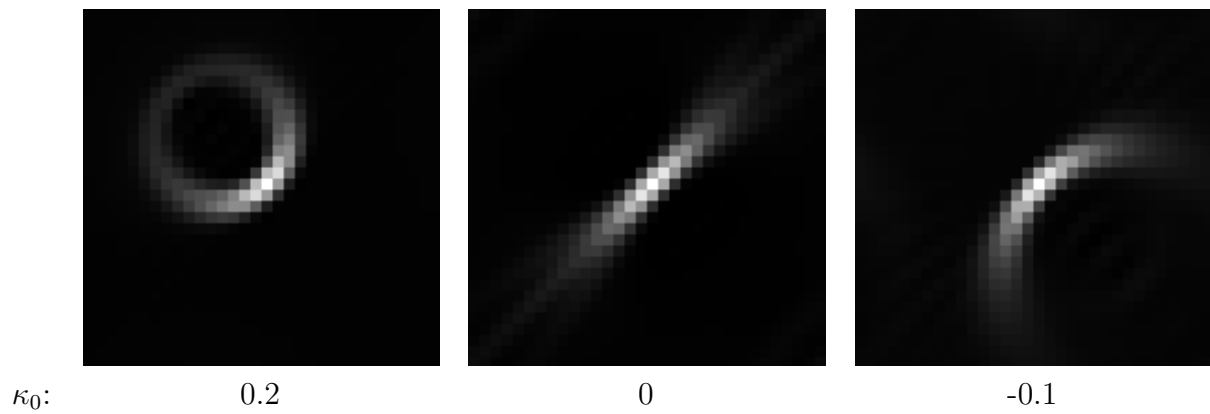


Figure 3.3: Impulse responses of the covariance matrix for the curve indicator random field for the curvature process. Impulses are located at the center of each image, directed at discrete direction 4 out of 32, with 5 curvatures. Parameters are $\sigma_{\kappa} = 0.01$, $\lambda = 10$. Each pixel represents an integral over all directions and curvatures associated with that position; the actual impulse response in each case is a sparse 4-dimensional array.

Chapter 4

Edge Correlations in Natural Images

Whether an algorithm ever works depends on the correctness of the assumptions upon which it is based. It is therefore useful to determine if the CIRF model of contours bears any resemblance to natural contour images so that our efforts are not wasted on an unrealistic model. Last chapter we derived the covariances of the CIRF U ; here we present empirical edge covariances for several natural images. The qualitative similarity between the covariances for our model and the observations suggests that the CIRF based on the direction process is a reasonable starting point for modeling image contours; the results also suggest a role for the curvature process.

Although we cannot measure the covariances of U directly (since we cannot observe its realizations), we can estimate covariances in local edge operator responses, or what we shall the measurement field M . These correlations¹ can reveal, albeit crudely, the qualitative structure of the correlations of U .

¹Following common practice, the terms correlation and covariance will be used interchangeably unless there is confusion.

While a detailed treatment of observation models will be presented in Chapter 5, we briefly mention the connection between the covariances Σ_M and Σ_U , of M and U , respectively. Suppose that the measurement field M was generated by the corruption of the CIRF U with blur operator B with additive noise N , statistically independent of U . Because of the independence of N , the cumulant generating functional of the measurements $M = BU + N$ is the sum of the cumulant generating functions for BU and N respectively. This implies that the measurement covariance is $\Sigma_M = \Sigma_{BU} + \Sigma_N = B\Sigma_U B^* + \Sigma_N$, where B^* is the transpose of B . If the noise is relatively white and the blur has relatively small scale, then this result says that the measurement covariances will be a slightly blurry form of the CIRF covariances, but different along the diagonal. Therefore studying the measurement covariances is warranted for verifying the plausibility of the CIRF.

To carry out our observations, we use oriented edge filter banks. Specifically, at each image position (x, y) and direction θ we shall measure the edge strength $m(x, y, \theta)$, which is a realization of the random variable $M(x, y, \theta)$.² The resultant set of these random variables forms our measurement random field $M = M(\cdot)$. The analysis in the previous paragraph suggests that for good continuation to have any meaning for natural contour images, at least the following must hold:

Conjecture 1. *For natural images, M is not “white,” i.e., its covariance*

$$\Sigma_M(r_1, r_2) := E[(M(r_1) - E[M(r_1)])(M(r_2) - E[M(r_2)])] \neq 0,$$

for all $r_1 \neq r_2$, where $r_i := (x_i, y_i, \theta_i)$.

²In this chapter we are using explicit spatial coordinates instead of the abstract site index i .

In §4.2, we shall present direct evidence not only in support of this conjecture, but also in agreement with the CIRF model: the observed edge correlations have an elongated pattern.

After first presenting the results in this chapter [5], we found a paper by Okajima [88] in which edge correlations of natural images were observed and used for learning edge filters. After publishing our work [3], we found that Kruger [64] and Schaaf [123] had already independently observed an elongated structure in edge correlations as well. This research area has recently become very active, with edge correlation reports in [34] and [110]. One distinction of our work is that we assume the statistical homogeneity (§4.1) that is qualitatively apparent in others' results, allowing us to compute edge correlations from a single image. This suggests that it is possible to adapt or “tune” contour filter parameters to individual images. A more important difference is the theoretical CIRF framework we provide for interpreting these observations.

More generally, Elder and Krupnik have recently estimated the statistical power of grouping cues other than good continuation [25]. Martin et al have computed statistics of Gestalt cues as well, but using hand-segmented images instead [74]. In related work on image statistics, the joint distributions of neighboring wavelet coefficients have been measured [111, 44]. Mumford and Gidas have recently suggested an axiomatic generative model of natural images using infinitely divisible distributions [83]. We currently ignore higher-order moments in order to focus on the longer distance spatial variation of edge relations. Invoking techniques from the study of turbulence, self-similarity properties in image gradients have also been measured [120]. For a review of natural images statistics, see [104].

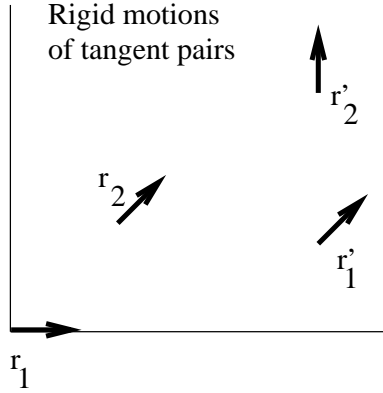


Figure 4.1: If the measurement random field is homogeneous, then the correlation of the (edge) measurements located at the primed pair of tangents equals that for the unprimed pair.

4.1 Statistical Homogeneity

for Random Fields in $\mathbb{R}^2 \times \mathbb{S}$

Unfortunately, since $\mathbb{E}[M(x_1, y_1, \theta_1)M(x_2, y_2, \theta_2)]$ is a function of six variables, it is clear that computing, storing, and visualizing this object—and thus assessing Conjecture 1—would be difficult without simplifications. Inspired by the Euclidean invariance of stochastic completion fields [126], we suggest that the proper analogue to the translation homogeneity of (x, y) -images for (x, y, θ) -space is the following translation and rotation homogeneity in edge correlations (Fig. 4.1):

$$\mathbb{E}[M(x_1, y_1, \theta_1)M(x_2, y_2, \theta_2)] = \mathbb{E}[M(0, 0, 0)M(\text{rot}_{-\theta_1}[x_2 - x_1, y_2 - y_1]^T, \theta_2 - \theta_1)],$$

$$\text{where } \text{rot}_\phi = \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix}.$$

This notion of homogeneity in correlation can be naturally extended to all statistics of random fields on $\mathbb{R}^2 \times \mathbb{S}$.

4.2 Empirical Edge Correlations

To test Conjecture 1, we measured edges using the logical/linear edge operator [50]. Here and throughout this thesis the default scales of $\sigma_{\text{tangential}} = 2.5$ pixels (length), $\sigma_{\text{normal}} = 1.4$ pixels (width) were used. The correlations were then estimated via: $\text{const} \cdot \Sigma M(r_1)M(r_2)$, $\forall r_1, r_2$ such that $(\text{rot}_{-\theta_1}[x_2 - x_1, y_2 - y_1]^T, \theta_2 - \theta_1) = (x, y, \theta)$. An appropriate spatial normalization was used to account for the relative lack of distant tangent (edge) pairs over nearby pairs. The results show first (Fig. 4.2) that natural images have a more extended correlation pattern than does pure noise, supporting the Conjecture. Other experiments (Fig. 4.3) show a similar correlation pattern. They also bear a qualitative resemblance to the correlations in the curve indicator random field (Fig. 3.2), which suggests that the orientation-based curve organization models [80, 139, 39] may have a statistical basis.

To more fully compare the distance over which edges are correlated, we plotted the correlation pattern along the central horizontal strips in the 0-orientation slices of the correlations (Fig. 4.4). The most important observation is that collinear edges are correlated over large distances. More surprising is the fact that the actual length of the contours in the blood cells image (the true perimeter of the cells is over 100 pixels) is not reflected in its collinear edge correlations. That there is statistical structure to be exploited in this image is shown in Fig. 4.5, which reveals a curved pattern in the distribution of correlations. These observations not only suggest that the above straight-line-seeking direction-based curve organization models would be ineffective on this image, but they point to a way out. By including curvature as an explicit variable in a curve organization framework [138], one may more fully capture the local geometry of the highly curved contours. The curvature process (§2.4) provides a random model for these observations. With deeper confidence about the

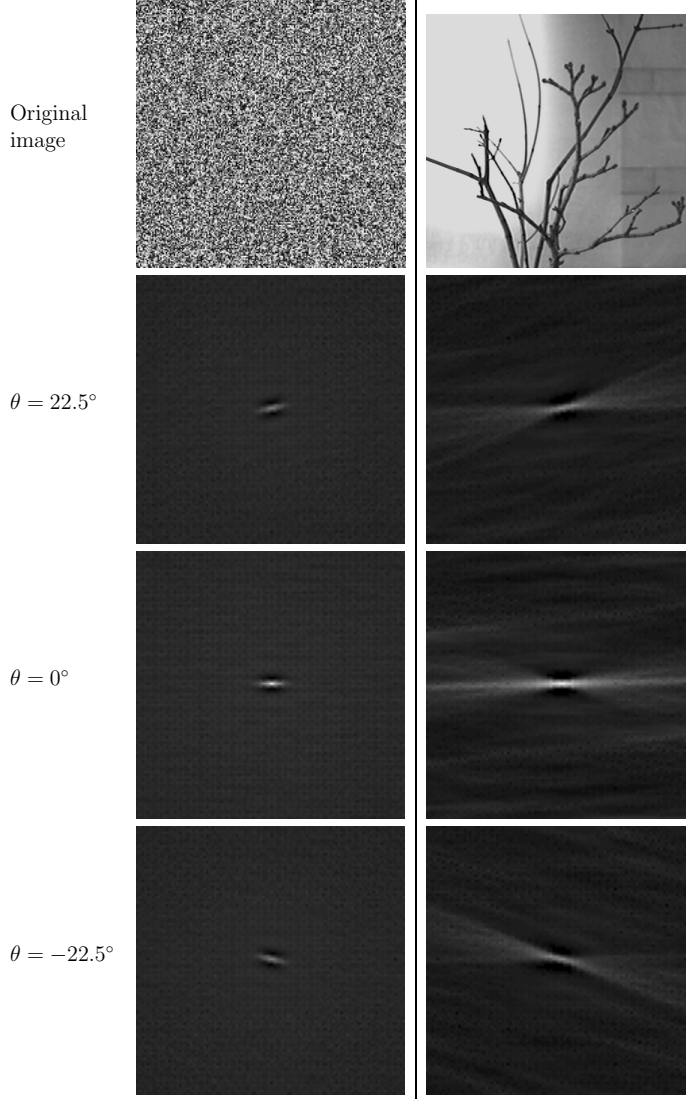


Figure 4.2: The empirical edge covariance function $\mathbb{E}[M(0, 0, 0)M(x, y, \theta)]$ of a pure noise (i.i.d. uniform(0,255)) image (top left) is contrasted with that of a natural image with contours (top right). Constant- θ slices of the covariances are shown for $\theta = 22.5^\circ, 0^\circ$, and -22.5° in the bottom three rows, where whiteness indicates amount of covariance. (These correlation images, each 100×100 pixels, have been enhanced by setting the gamma correction to 2.) The slight edge correlations in the noise image (left) are due entirely to the support of the logical/linear edge operator. The elongation on the right indicates that edges are strongly correlated in the tangential direction. Note the resemblance to the covariances of the direction process-based CIRF (Fig. 3.2). We consider this as empirical evidence of good continuation and proximity in a natural image. Again, the pattern of correlations is also consistent with the models in [138, 39, 126].

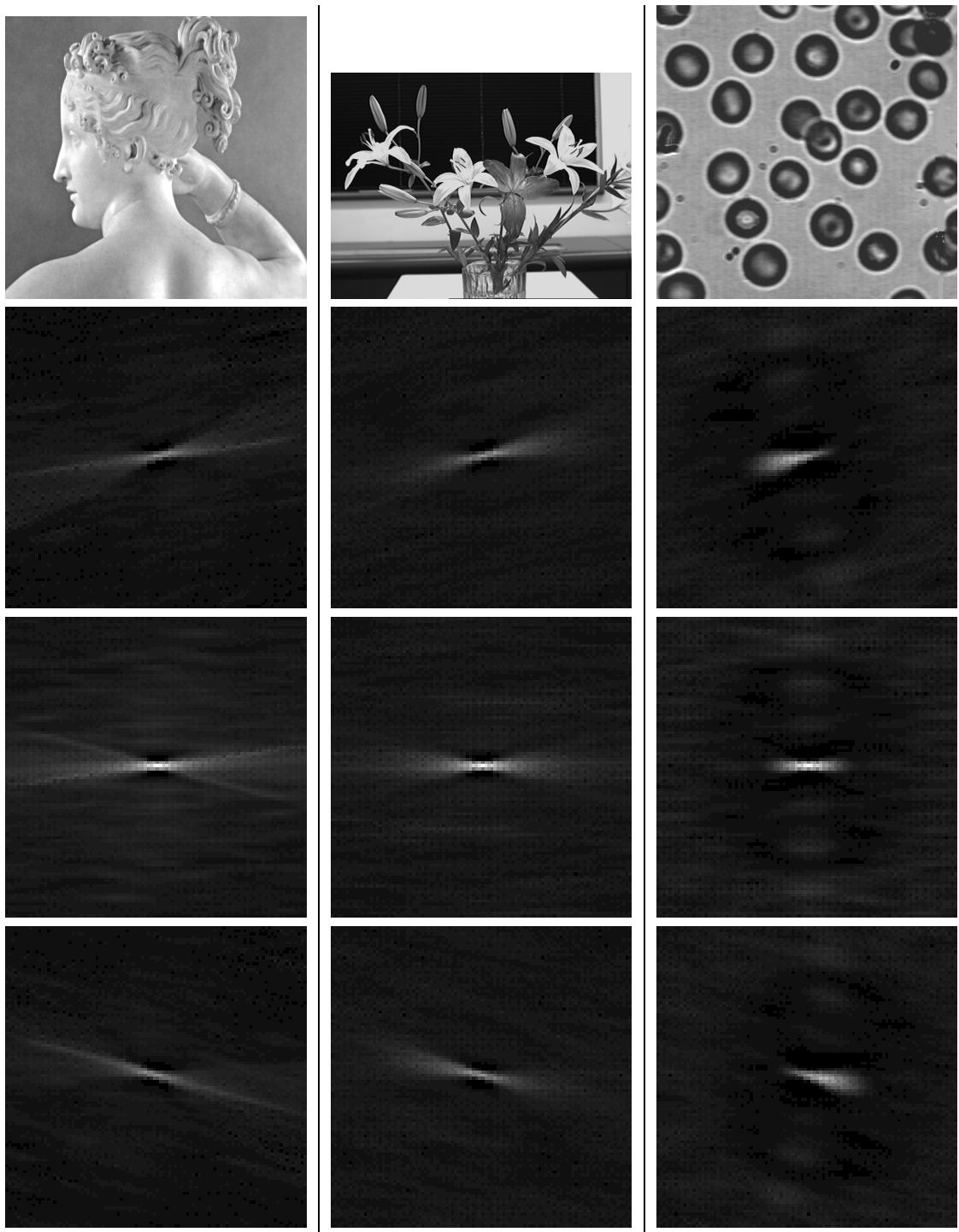


Figure 4.3: Edge covariances (continued; see caption of Fig. 4.2 for explanation). The Paolina (left) and lily (center) images have correlations in the tangential direction which extend much farther than those for the blood cells image (right). Note how these more distant correlations in the Paolina image occur despite the contours being open, unlike the closed contours in the blood cell image, suggesting curve enhancement systems should not rely only on contour closure.

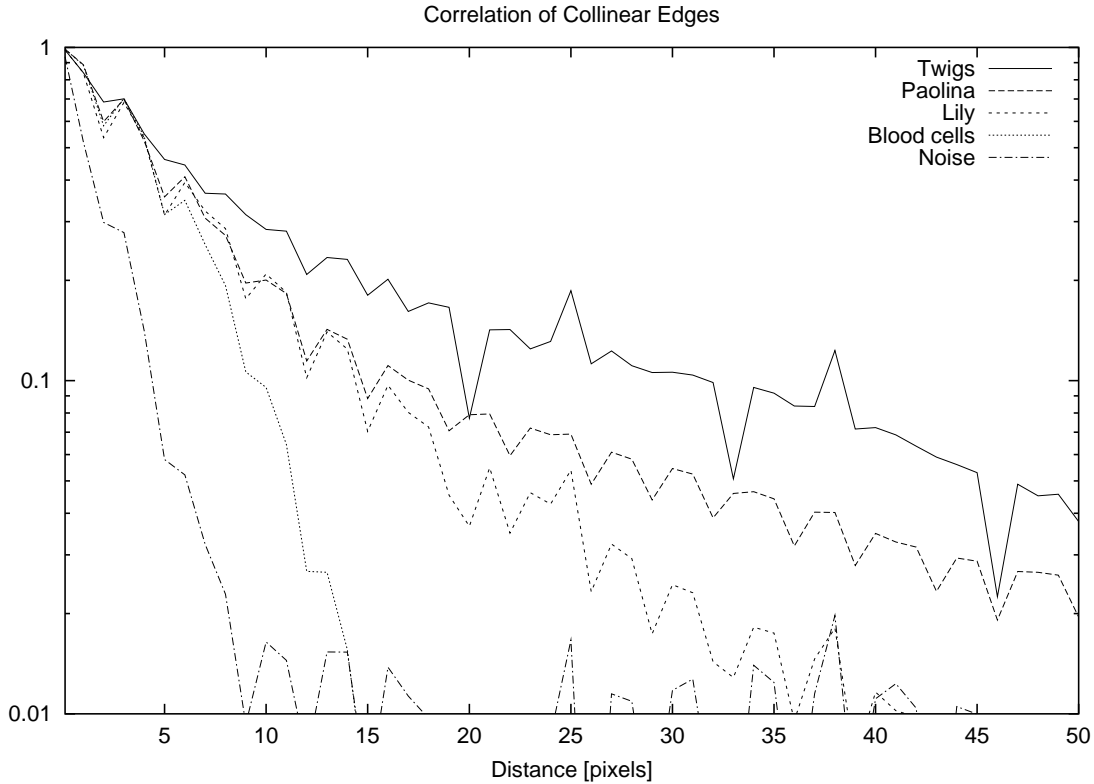


Figure 4.4: Comparison of the correlation coefficient between collinear edges, obtained by extracting a central strip $\mathbb{E}M(0, 0, 0)M(x, 0, 0)$ through the zero-orientation slices of the correlations in Figs. 4.2 and 4.3, subtracting the square of the mean (estimated by taking a spatial average in an approximately constant region of the covariance), and then dividing by the estimated variance of M . Observe that although the correlation of a pure noise image falls off rapidly, the correlation for natural images persists over much longer distances. For example, note that even at 25 pixels of separation, collinear edges in the twigs (Fig. 4.2, right), Paolina and lily (Fig. 4.3, left and center) images are appreciably more correlated than pure noise. The correlation of the blood cells image drops off rapidly due to high curvature (see Fig. 4.5).

relevance of the CIRF to real images, we now turn to the task of deriving contour enhancement filters.

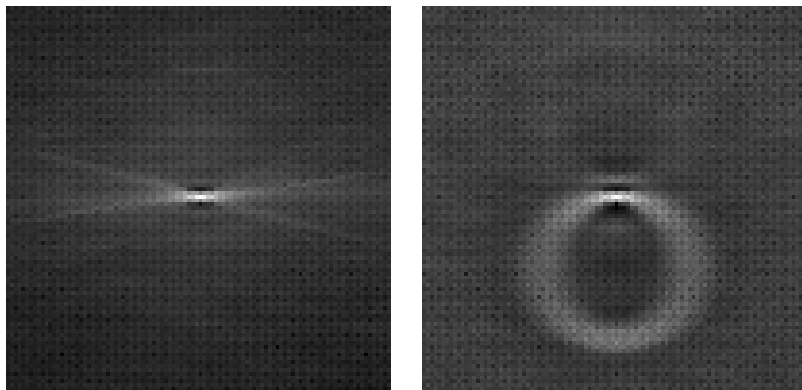


Figure 4.5: Edge correlations as a function of position only. The orientation integral $\int_0^{2\pi} \mathbb{E}[M(0,0,0)M(x,y,\theta)]d\theta$ for the Paolina and blood cell images (left and right, respectively, of Fig. 4.3) are shown (left and right, respectively). Observe how the positive “ridge” of the correlation bends around in a circle (right), just as the image of blood cells is composed of circles. The black-blood-cell/white-background asymmetry is captured in the vertical asymmetry of the correlation. More importantly, this example shows that the apparently short-distance correlations of the blood cells (Fig. 4.4) were observed because we only considered collinear edges there. It is the non-collinear, indeed the co-circular [89] edges that give rise to the circular correlation pattern (right). Thus this figure supports the use of curvature in curve enhancement systems [138] to exploit longer-distance correlations: the contour structure to be enhanced may exist along a straight line (left) or along a circle (right).

Chapter 5

Filtering Corrupted CIRFs

In this chapter we focus on the central objective of this thesis:

Suppose an unknown number of contours are located within a set of sites \mathcal{I} . Consider the question \mathcal{Q}_i : “Did a curve pass through location i ?”

Task: *Given a corrupted image of the curves in \mathcal{I} , compute the optimal answer to question \mathcal{Q}_i , for all i in \mathcal{I} .*

Unfortunately, this task definition leaves a number of crucial questions unanswered, such as:

1. What is an image of contours?
2. How is the curve indicator field related to the $\{\mathcal{Q}_i\}$?
3. What do we mean by corruption?
4. What do we mean by “optimal”? Is this notion unique? If not, which optimal is optimal?

5. What is the optimal answer? Is it computable? If so, is the computation tractable?

In this chapter we first consider the conceptual issues raised by questions (1) through (4) above, leading to a clearer formal specification of what will become our filtering Task. We then handle question (5) by providing derivations of several linear and nonlinear filters that approximately satisfy our exact specification of filtering noisy contour images.

Before addressing these issues, let us consider whether it is worth the effort to perform this Task at all. Certainly boundary representations of our natural world have been with us for a long time: even early cave drawings exhibited clear contour structure, and artists have obsessed over lines ever since. In computer vision and image processing, researchers have been interested in curves to support a host of computations, including object recognition, tracking, and image compression. However, the natural task would seem to be that of estimating the curves themselves, not whether a curve had passed through a given point. Because we shall return to this issue in §5.9, at this point we only mention that successfully performing our Task supports the computation of the curves themselves via an edge-linking procedure.

On the surface, satisfying our objective appears startlingly simple: it seems one should simply take one pass over the image, where for each site i one makes a decision, perhaps based on a pixel value or local edge operator output. Indeed, this two-step procedure of local filtering followed by thresholding has been a staple of computer vision for almost four decades, beginning with Roberts in 1965 [100]. Unfortunately, in a wide variety of situations the results of this procedure are unsatisfactory, and a number of attempts at more sophisticated computations have proved insufficient to sway the field into depending upon low-level contour information. This thesis attempts to tackle this contour finding problem at the root by defining the problem

exactly in a well-understood probabilistic inference framework, and then systematically deriving procedures for (approximately) solving it.

We begin by relating the questions $\{\mathcal{Q}_i\}$ to the curve indicator random field. Because the correct answer to \mathcal{Q}_i is “yes” or “no,” and because we are uncertain about the correct answer, we can associate a binary random variable with site i , whose value will be 1 if a curve passed through site i and 0 if not. Recalling that the curve indicator random field U_i is the total amount of time spent at i by any curve, the correct answer to \mathcal{Q}_i can be exactly represented by the *binary curve indicator random field* $U_i^{>0} := \mathbb{1}\{U_i > 0\}$, where i is in \mathcal{I} . Notice that $U^{>0}$ captures a complete notion of what we mean by an ideal edge or line map, or what we shall call a (perfect) image of curves, or contour image.¹ Let the measurement vector m denote the given corrupted contour image. Because we are uncertain about the exact pattern of corruption, it is natural to view m as a realization of a measurement random field M . Therefore our Task is one of determining $U^{>0}$, given the measurements m .

5.1 Bayes Decision Theory

Realize that our Task mandates that we make a particular *decision* about $U^{>0}$, i.e., that we form some estimate $\widetilde{u^{>0}}$ of $U^{>0}$. Because $U^{>0}$ and M are random, our Task therefore falls under the purview of Bayes decision theory, which we review now. For an introduction to decision theory, see [13, Ch. 10] or [20, Ch. 1].

Suppose we seek to make a decision \mathcal{D} about an unknown state of nature $X = x \in \mathcal{X}$ distributed with prior $p(x)$, given only observation $Y = y$ (conditioned upon the unknown x) distributed with likelihood $p(y|x)$. The loss function $\text{loss}(x, \mathcal{D})$ specifies the penalty incurred by making decision \mathcal{D} given that the state of nature was in fact

¹To relate to the work of Parent and Zucker, we note that since $U^{>0}$ is also an indicator function over the union of the traces of the curves, and so $U^{>0}$ could also be called a trace indicator function.

x . Because this decision is based upon the observations, we know \mathcal{D} is a mapping associating the decision $\mathcal{D}(y)$ to the observation y . We focus on that class of decision problems where the decision \mathcal{D} is an *estimate* of the unknown x . The goal in Bayes decision theory is to select that (Bayes) decision rule or estimator $\tilde{x} = \tilde{x}(y)$ which minimizes the expected loss given y , or

$$\tilde{x}(y) := \arg \min_x \mathbb{E}[\text{loss}(X, x)|y]. \quad (5.1)$$

The Bayes risk $\mathbb{E}[\text{loss}(X, \tilde{x}(Y))]$ is the overall expected loss (averaged over all observations Y) for the Bayes estimator.

5.1.1 Previous Estimation Problems

Bayesian and related techniques for estimation have a long history in signal process, image processing, and computer vision. These methods emphasize the tension between fidelity to observation (as defined by the likelihood) and some assumed structure or regularity in the unknown (as defined by the prior). Early work was with 1-dimensional signals for communications [77] applications. The Wiener filter exploited knowledge of signal and noise correlations to minimize mean squared error [124, 94]. Tracking applications [113] are more appropriate for the Kalman-Bucy filter [94, 14]. Here, the unknown signal (a function of time) is modeled by a (possibly time-varying) linear stochastic system (with finite dimensional state-space) and the observations have additive Gaussian noise; the filter provides an optimal estimate of the state of the system given the past observations. Cox et al even applied a form of the Kalman filter to contour tracking in images [15], where the stochastic processes are the individual curves. Somewhat surprisingly, the Kalman filter has been applied to image restoration problems, where the image is arranged into a time

series by concatenating successive rows of pixels [132, 131, 8]. Hidden Markov models (HMMs) generalize the Kalman-Bucy filter to handle nonlinear systems by allowing (general) Markov process signal models (whose state is hidden) and general (state-dependent) observation models [27]; a major application of HMMs has been speech recognition [98, 97]. Particle filters, which approximate the distribution over the unknown current state of the process with a large number of particles, have been used for visual tracking [48] and mobile robot localization [119]. In binocular stereopsis, Markov processes have also been used to model the unknown disparity signal, where the observation comes from a stereo image pair [6].

Extending these statistical techniques to general image processing applications requires making the transition from random processes to random fields. For additive Gaussian observations with a Gaussian prior, the extension from the 1-dimensional case to 2-dimensional image restoration applications is straightforward [2, 7]. An interesting application of Bayesian methods to the restoration of CCD images of galaxies is [79], where observation nonlinearities are explicitly handled. More generally, Markov random fields (MRFs) have been used to capture local image interactions for many image processing and vision applications [130, 36, 128, 67]. While early applications required hand-tuning of parameters the MRF, recent work has been more data-driven, with learning procedures such as minimax entropy [137]. Energy minimization, or physics-based, approaches to computer vision are closely related to Bayesian methods²; applications include the reconstruction of images with discontinuities [9] and the local optimization of a contour to image information [54]. Clearly, estimation has important place in signal processing, image processing, and computer vision.

²Energy $E(x)$ can be converted to a probability $P(x)$ using $P(x) = Z^{-1} \exp(-E(x))$, where $Z := \int \exp(-E(x)) dx$.

To relate the Bayes decision framework to our Task, we note that the unknown state of nature X above is our contour image $U^{>0}$ and the observation Y above is the measurement random field M . This answers two of the questions we began with in this chapter. First, corruption of an image of contours is simply the random mapping taking $U^{>0}$ to M , and is completely characterized by the likelihood $p(m|u^{>0})$. Second, Bayes decision theory does not provide a unique notion of optimality, because the loss function must be specified. We shall return to this latter issue in §5.1.2.

The requirement of making a decision sets Bayes decision theory apart from related work such as “pure Bayesianism,” which merely reports the posterior distribution $p(x|y)$ without making a decision. Decision theory is similar to game theory both in that an action will be taken and that the loss function is qualitatively comparable to the payoffs; the difference lies in the notions of randomness employed. In game theory the task is often to determine the equilibria, or marginal distributions over the unknowns $X_i, i = 1, \dots, n$; the observations Y can be used to bias toward particular strategies via strategy-specific payoffs. In decision theory, the observations are included via the likelihood.

The likelihood is the centerpiece of inferences made by the classical statistical school, where the existence of the (Bayesian) prior is denied. In the context of edge detection in computer vision, this amounts to an assertion that all binary functions of i are equally preferable—a uniform prior on edge maps—; only the likelihood and loss function can be used to make the decision. If the measurements are local and conditionally independent (i.e., $p(m|u^{>0}) = \prod_i p(m_i|u_i^{>0})$), which is a common assumption, and if the loss is additive (see §5.1.2), then one is forced to do sitewise thresholding. The primary difficulty with this uniform prior is that it completely ignores the rich structure of images of curves. Not only are contour images often sparse, but their non-zero regions have a stereotyped behaviour: they are curves!

Even the relatively simple Markov process-based CIRF model of contour images is highly non-uniform, even non-Gaussian, as we saw in Chapter 3. The application of Bayes decision theory to our contour Task allows us to consistently exploit this known structure, to hallucinate contour-ness into images to overcome the ambiguities introduced by our imperfect measurements. Gestalt psychologists noted this idea qualitatively in the principle of good continuation [61, 52]. Even earlier Helmholtz suggested that the visual system is performing “unconscious inference,” where our perceptions are biased towards expected structure in the world.

5.1.2 Loss Functions: Which Best is Best?

Given this background, we now proceed to apply Bayes estimation for optimally determining $U^{>0}$. However, we still need to select a loss function to make our answer unique. It is well-known that the loss function can significantly influence an estimate when the posterior is not peaked, so we shall take some care in our choice. But because we are seeking a binary random field, our loss function is limited to the set of mappings from $\{0, 1\}^{|I|} \times \{0, 1\}^{|I|}$ to \mathbb{R} . Even that set of possibilities is huge, so we focus on two opposite extremes: the 0/1 and the (symmetric, uniform) additive losses.

The 0/1 loss assigns a penalty of 1 if the estimate is different from the true unknown, and 0 only if the estimate is correct. It is a standard exercise to show that the Bayes estimate under a 0/1 loss is exactly the maximum a posteriori (MAP) estimate, or posterior mode. Estimating an unknown by choosing the most probable possibility is so intuitive and so common that many applications of Bayes rule are (almost) synonymous with MAP estimation. Since the Bayes risk of the MAP estimate is the probability of error, the appeal of MAP estimation is obvious: who would want more error than the minimum?



Figure 5.1: The ground truth image on the left is mis-estimated by all candidates to its right. According to the 0/1 loss function, all of these mis-estimates are equally poor. In contrast, an additive loss function monotonically increases with *site-level* error rate.

Perhaps in an attempt to minimize error, MAP estimation has been common in image analysis, where it is often coupled with a Markov random field posterior. One author of a recent monograph on MRFs [67] even declared that MAP “has been the most popular choice in MRF modeling for image analysis” and calls this combination of MAP and MRFs “the MAP-MRF framework” [68].

Unfortunately, the basic property that the 0/1 loss function underlying MAP estimation penalizes all errors equally may be inappropriate for many image analysis applications, especially our contour Task. For example, whether the estimate was incorrect at only one site or at half of the sites would not influence the penalty (Fig. 5.1). What would be preferable is a measure that gracefully degrades with the difficulty of the task.

Symmetric, uniform, additive loss functions take a sum of site-wise errors (for binary images) and therefore their Bayes risk is the minimum average number of site-wise errors. Unlike the 0/1 loss, this site-wise error count does not abruptly change with any error at all, rather it gradually increases as the estimate becomes poorer. Perhaps it is the existence of two notions of “error rate” that has helped maintain the popularity of MAP in modern image analysis applications, long after early successes in pattern recognition. On the one hand there is the probability of (an) error (0/1 loss); on the other, there is the probable number of errors, or probable

error rate (additive loss). At low noise levels, the difference between additive and 0/1 losses may be small. At high noise levels, however, Marroquin observed empirically that MAP estimates of binary fields are particularly poor compared to estimates based on an additive loss [71].

5.1.3 Minimum Mean Squared Error Filtering

For binary image problems, this additive loss function is equivalent to the squared-error loss, or the sum of the squares of the differences between the true unknown and the estimate. Thus minimizing the additive loss amounts to choosing that binary field that minimizes the mean squared error (MMSE), or to computing

$$\widetilde{u^{>0}} = \arg \min_{b \in \{0,1\}^{|\mathcal{I}|}} \mathbb{E}_m ||U^{>0} - b||^2 \quad (5.2)$$

$$\begin{aligned} &= \arg \min_{b \in \{0,1\}^{|\mathcal{I}|}} \sum_i \mathbb{E}_m (U_i^{>0} - b_i)^2 \\ &= (\arg \min_{b_1 \in \{0,1\}} \mathbb{E}_m (U_1^{>0} - b_1)^2, \dots, \arg \min_{b_{|\mathcal{I}|} \in \{0,1\}} \mathbb{E}_m (U_{|\mathcal{I}|}^{>0} - b_{|\mathcal{I}|})^2), \end{aligned} \quad (5.3)$$

where the last line follows from the second because the terms in the sum of positive numbers can be optimized independently. Thus the additive loss provides an added benefit in complexity: we have a separation of the joint estimation problem over $2^{|\mathcal{I}|}$ possibilities into $|\mathcal{I}|$ single-site problems of two possibilities each. Also, the $|\mathcal{I}|$ separate expectations are taken with respect to the *marginal* distributions $p(u_i^{>0}|m)$, not the complicated joint distribution. In real-valued problems, one can expect the marginals to be better behaved (smoother) than the joint, since each marginalization is an integration process. (In this binary problem, the marginals are so “smooth” they are a function of a single number!) Thus MMSE estimation may not only be more relevant to image analysis than MAP estimation, it may be easier as well.

5.1.4 Bounding the MMSE Estimate of the Binary CIRF with a Thresholded Posterior Mean of the CIRF

Now that we have refined our original Task to that of computing the MMSE estimate $\widetilde{u}^{>0}(m)$ of the binary CIRF $U^{>0}$ given measurements m , we would like to articulate this goal in terms of the (non-binary) CIRF U , for which we have a complete and tractable statistical characterization.³ While it is obvious that $U^{>0}$ and U are related, the extent to which knowledge about U can help us in estimating $U^{>0}$ is not. In this section we show that $\widetilde{u}^{>0}(m)$ and a thresholding of $\mathbb{E}_m U$ are intimately related by a simple inequality.

Because the squared-error loss is additive, we saw in (5.3) that we can focus on the problem of finding the MMSE estimate $\widetilde{u}^{>0}(m)$ at site i , which can be expressed as

$$\begin{aligned}
\widetilde{u}^{>0}_i(m) &= \arg \min_{b_i \in \{0,1\}} \mathbb{E}_m (U_i^{>0} - b_i)^2 \\
&= \arg \min_{b_i \in \{0,1\}} \mathbb{E}_m (U_i^{>0})^2 - 2b_i \mathbb{E}_m U^{>0} + b_i^2 \\
&= \arg \min_{b_i \in \{0,1\}} (\mathbb{E}_m U_i^{>0} - b_i)^2 \\
&= \begin{cases} 1, & \text{if } \mathbb{E}_m U_i^{>0} \geq \frac{1}{2} \\ 0, & \text{else} \end{cases} \\
&=: \mathbb{1}\{\mathbb{E}_m U_i^{>0} \geq 1/2\}.
\end{aligned} \tag{5.4}$$

Thus the MMSE estimate of our binary field is its thresholded posterior mean [71], suggesting that we first seek to relate the posterior means of $U_i^{>0}$ and U_i , and then threshold.

³The original intent was to exploit the statistics of $U^{>0}$ directly. The CIRF is used instead because stronger results could be derived.

Before pursuing that relationship, we first study an intermediary, the *hitting random field* H , where $H_i \in \{0, 1, 2, \dots\}$ is the total number times that site i was hit by any curve Markov process. Observe that the binary CIRF is a simple thresholding of the hitting random field, i.e.,

$$U_i^{>0} = \mathbb{1}\{H_i > 0\}, \quad (5.5)$$

and so any function of $U^{>0}$ is also a function of H . In particular, while it might appear most natural to assume that our measurements are a corrupted version of the binary CIRF $U^{>0}$, there is no loss in generality in assuming that our measurement random field M is a function $M(H, N)$, where N is some noise vector independent of the curves.

Let τ_i denote the average amount of time that the curve Markov process spends in site i each time it hits i (without conditioning on m). It can be shown that $\tau_i = -1/q_{ii}$, where $Q = (q_{ij})$ is the generator of the killed Markov process [22] (if the Markov processes is spatially homogeneous (see §5.6.3), then τ_i does not vary with i). The following lemma relates the posterior means of the curve indicator and hitting random fields by the scalar factor τ_i .

Lemma 3. *Suppose that m is a realization of the measurement random field $M = M(H, N)$, where noise N is independent of the curve Markov processes (and thus the hitting random field H and the CIRF U). Then:*

$$\mathbb{E}_m U_i = \tau_i \mathbb{E}_m H_i, \quad \forall i \in \mathcal{I},$$

where τ_i is the average time spent in site i per visit.

PROOF. Let $t_i^{(n)}, n = 1, \dots, H_i$ denote the (random) holding time spent in site i during the n -th hit by a curve Markov process. (The ordering of the hits is immaterial; the index n merely distinguishes the different time intervals.) Recall from the theory of continuous-time, discrete-state Markov processes [86] that the holding times are independent, exponentially-distributed random variables of mean τ_i , when conditioned upon the sequence of states taken by the processes. It follows that the $t_i^{(n)}$ ($n = 1, \dots, H_i$) are independent exponentially-distributed random variables of mean τ_i , given H . We can write the CIRF at site i as $U_i = \sum_{n=1}^{H_i} t_i^{(n)}$, and therefore $\mathbb{E}_M U_i = \mathbb{E}[U_i|M] = \mathbb{E}[\sum_{n=1}^{H_i} t_i^{(n)}|M] = \mathbb{E}[\mathbb{E}[\sum_{n=1}^{H_i} t_i^{(n)}|H, M]|M] = \mathbb{E}[\sum_{n=1}^{H_i} \mathbb{E}[t_i^{(n)}|H, M]|M]$. We now observe that $\mathbb{E}[t_i^{(n)}|H, M] = \mathbb{E}[t_i^{(n)}|H] = \tau_i$, where $t_i^{(n)}$ given H is independent of M because (1) M is only a function of H and N , and (2) N is independent of $t_i^{(n)}$ (because it is determined by the curve Markov processes, which are independent of N). The result follows because $\sum_{n=1}^{H_i} \tau_i = \tau_i H_i$. \square

Given this connection between U and H , we are now in a position to show that the thresholded posterior mean of the CIRF U is an upper bound on the MMSE estimate $\widetilde{u}^{>0}$ of the binary curve random field $U^{>0}$.

Proposition 6. *Suppose that m is a realization of the measurement random field $M = M(H, N)$, where noise N is independent of the curve Markov processes. Then the MMSE estimate of $U^{>0}$ given m is bounded from above by a thresholded posterior mean of the CIRF, or*

$$\widetilde{u}^{>0}_i(m) = \mathbb{1} \{ \mathbb{E}_m U_i^{>0} \geq 1/2 \} \leq \mathbb{1} \{ \mathbb{E}_m U_i \geq \tau_i/2 \}, \quad \forall i \in \mathcal{I}, \quad (5.6)$$

where τ_i is the average time spent in site i per visit.

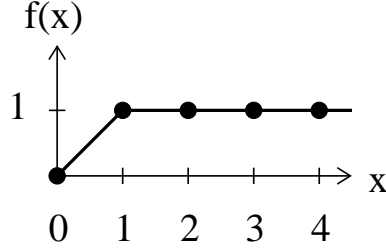


Figure 5.2: Function $f(x)$ used in the proof of Proposition 6.

PROOF. There are two steps to the proof. First we construct a concave function f that will enable us to put an upper bound on the posterior mean of $U_i^{>0}$ in terms of the posterior mean of H_i . Then we threshold this inequality.

Consider the function

$$f(x) := \begin{cases} x, & \text{if } x < 1 \\ 1, & \text{if } x \geq 1. \end{cases}$$

See Fig. 5.2. Since $f(x) = \mathbb{1}\{x > 0\}$ when x is a nonnegative integer, we know that $U_i^{>0} = f(H_i)$. (Think of f , which is the piecewise linear graph in the figure, as an interpolator for $\mathbb{1}\{\cdot > 0\} : \mathbb{Z} \rightarrow \{0, 1\}$, which is indicated by the dots.) Observe that f is concave; therefore Jensen's inequality [13, p. 182–3] asserts that $\mathbb{E}_m f(H_i) \leq f(\mathbb{E}_m H_i)$. Using Lemma 3 we conclude that

$$\mathbb{E}_m U_i^{>0} \leq f(\mathbb{E}_m U_i / \tau_i). \quad (5.7)$$

Because the MMSE estimate $\widetilde{u_i^{>0}}$ is a thresholding of the posterior mean of $U_i^{>0}$ at $1/2$ (from (5.4)), we similarly threshold (5.7) at $1/2$. The result follows because the thresholding function $\mathbb{1}\{x \geq 1/2\}$ is a monotonic in x and because $\mathbb{1}\{f(x) \geq 1/2\} = \mathbb{1}\{x \geq 1/2\}$. \square

Proposition 6 enables us to connect the (non-binary) CIRF U to our binary decision problem. This upper bound has a simple interpretation: the thresholded posterior mean of the CIRF can have no more misses than $\widetilde{u}^{>0}$; it can only have more false positives. Because of this close relationship, in the rest of this thesis we shall approximate the MMSE estimate of $U^{>0}$ given m with the thresholded posterior mean of the CIRF U . Since thresholding is trivial, *we shall define our filtering problem to be the computation of the CIRF posterior mean $\mathbb{E}_m U$ only.* In statistical language, the posterior mean $\mathbb{E}_m U$ is known as the regression of U on m .

It is well-known that the posterior mean is the MMSE estimate for real-valued random variables (see [14, p. 68], [13], or [41, p. 483]), and therefore *our goal is equivalent to the computation of the MMSE estimate of U given m .* In many non-Gaussian cases the posterior mean can be difficult to compute; in our context, however, we are fortunate to have tractable formulas for the cumulants and moment generating functional to simplify our computations. Before developing linear and nonlinear MMSE CIRF filters that leverage these formulas, we shall present our observation model.

5.2 Observation Models and Likelihoods

Although the CIRF has been the focus of Chapters 2 and 3, it is only part of the contour inferencing story: the prior. Another key component is the likelihood, which we consider here.

5.2.1 The Gaussian Likelihood

The likelihood represents how our ideal contour image relates to the measurements, and therefore it characterizes the corruptions in local edge and line operator responses, or how observed contours are non-ideal. For example, edges are blurry, in

the sense that edge operators have an associated scale [129, 59, 26]. Equally important is the random noise in edge measurements: even for intensity images with high signal-to-noise ratio (SNR), edges can still be noisy for two reasons. Most fundamentally, edge detection requires differentiation, which amplifies noise. In addition, when the local contrast (i.e., signal) is low, the local SNR drops for fixed sensor noise.

One way of specifying such sources of corruption is to write an observation equation that relates the unknowns to the observations. Although one can consider much more general models, both for concreteness and simplicity we assume that the measurements are a linear function of the CIRF plus Gaussian noise N with zero-mean and covariance operator Σ_N , or $M = BU + N$, where $B \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{I}|}$ is a linear operator used to model the local edge or line filtering and blur. This observation equation gives rise to the following Gaussian likelihood that models noise and blur:

$$p(m|u) = (2\pi)^{-|\mathcal{I}|/2} (\det \Sigma_N)^{-1/2} \exp \left[-\frac{1}{2} \langle m - Bu, \Sigma_N^{-1} (m - Bu) \rangle \right].$$

We stress that the linear operator B is quite flexible: not only can it be a standard blur operator, but it can also represent more fundamental aspects of observation. For example, in the case where \mathcal{I} is the discretized three-dimensional (x, y, θ) -space and M is a two-dimensional image, B would be the (discrete) linear projection that integrates along θ . It can also capture the image derivatives used in local edge and line operators as well, and thus the operator B need not even perform blurring. Noise structure can be partially captured by non-diagonal noise covariance operator Σ_N , i.e., the correlated noise can be handled.

One may object with this likelihood model on the grounds that the measurement is a now function of the CIRF U , not the more natural binary CIRF $U^{>0}$ or even hitting random field H . Indeed, Proposition 6 required that $M = M(H, N)$, not $M =$

$M(U, N)$, and we are therefore introducing another approximation with this CIRF-dependent likelihood model. The key difference between U and H is an additional holding-time “noise” from the continuous-time, discrete-state Markov processes (see the proof of Lemma 3). At least for those cases where the kernel for B is positive and has large support (i.e., B is a large scale blur operator) and where $\tau_i = \tau$ for all i , the approximation $BU \approx \tau BH$ may be very good because the holding times are independent, have finite variance and therefore the central limit theorem applies. As the spatial discretization becomes finer but the absolute scale of the blur kernel remains fixed, this large (discrete) scale requirement will be met.

5.2.2 Log-Likelihoods with Quadratic Dependence on the CIRF

The Gaussian is an example of a more general class of *quadratic log-likelihoods* with

$$\ln p(m|u) = \epsilon \langle a + Au, u \rangle + \text{constant}, \quad (5.8)$$

where A is a symmetric, nonnegative definite matrix, and vector a and the constant are functions of the measurement vector m . To relate this class to the Gaussian likelihood (5.2.1), we set $\epsilon a = B^* \Sigma_N^{-1} m$ and $\epsilon A = -\frac{1}{2} B^* \Sigma_N^{-1} B$. The coefficient ϵ is introduced to capture the degree of noise; when the noise is white ($\Sigma_N = \sigma_N^2 I$), this coefficient could be used to represent the inverse noise variance, or $\epsilon = \sigma_N^{-2}$. Although for the Gaussian, the vector a is linear in the measurements, the quadratic log-likelihood allows a to be a general functions of m . Indeed, all of the nonlinear filters we shall be presenting allow a to be any (measurable) function of m . Some non-Gaussian observation models can be handled using this freedom in a and A by a change of variables to a Gaussian model. For example, multiplicative noise can be

modeled using a logarithmic change of variables to additive Gaussian white noise. Molina et al in effect used this technique for restoring astronomical images [79, Equation (24)]. We present another non-Gaussian example of the quadratic log-likelihood next.

5.2.3 Exploiting Empirical Distributions of Local Operators

Suppose we have training data in the form of a number of curve images and their corrupted counterparts. How could we use information gathered from this data to empirically derive our likelihood function?

One way of doing this is to construct an observation model that only considers whether a contour hit a site or not by determining the “on” and “off” edge responses [35]. To formalize this, we set up a likelihood in terms of the binary CIRF $U^{>0}$, where $U_i^{>0} = 1$ only when at least one of the contours passed sometime through site i , and 0 otherwise. To simplify the situation, we focus on the class of conditionally independent, local likelihoods: $p(m|u^{>0}) = p(m_1|u_1^{>0}) \cdots p(m_{|I|}|u_{|I|}^{>0})$. Following [35, 133], we consider two distributions over measurements at site i : the “on” distribution $p_{\text{on}}(m_i) := p(m_i|u_i^{>0} = 1)$ and “off” distribution $p_{\text{off}}(m_i) := p(m_i|u_i^{>0} = 0)$. (Spatial homogeneity is implicit here: the “on” and “off” distributions are assumed not to depend on i , although this can be relaxed.) Such distributions can be approximated with histograms obtained from the training data [63]. It follows that $\ln p(m|u^{>0}) = \sum_i u_i^{>0} \ln \frac{p_{\text{on}}(m_i)}{p_{\text{off}}(m_i)} + \sum_i \ln p_{\text{off}}(m_i)$. Letting τ again be the average amount of time that the Markov process spends in a site per hit of the site, observe that U_i/τ and $U_i^{>0}$ are equal on average, *as long as each site gets hit at most once*. Ignoring this latter condition for a moment, one can heuristically try to create a likelihood conditioned on u by replacing $u^{>0}$ with u/τ above: $\ln p(m|u) \approx \epsilon \sum_i a_i u_i + \text{constant}(m) = \epsilon \langle a, u \rangle + \text{constant}(m)$, where $a_i = a_i(m_i) := \tau^{-1} \ln \frac{p_{\text{on}}(m_i)}{p_{\text{off}}(m_i)}$

and $\epsilon = 1$. To ensure normalizability, we include a term $\epsilon \langle Au, u \rangle$ in the likelihood, where $A = -\frac{1}{2}I$, and so obtain a valid quadratic log-likelihood in the form of (5.8). Observe that this log-likelihood has quadratic dependence on u , but possibly nonlinear dependence on m . The norm-squared $\|u\|^2 = -2\langle Au, u \rangle$ measures the self-interaction of the CIRF; specifically, this term is $\sum_i u_i^2$, or the sum of the measure of the self-intersections and crossings of the curves as described in §3.1.1. This quadratic term therefore penalizes violations of the above condition that a site gets hit at most once. (To further penalize intersections, one can replace the old prior $P(du)$ with a modified prior $Z_\beta^{-1} \exp(-\beta\|u\|^2)P(du)$, where $Z_\beta := \int \exp(-\beta\|u\|^2)P(du)$. See §5.3 for notation.) This empirical model can also be combined with the 3-d to 2-d projection discussed in the Gaussian case by modifying A appropriately if the training observations are two-dimensional.

5.3 The Posterior Distribution

In this section we set up some notation for gluing together the likelihood with the prior to form the posterior. We write the CIRF prior as $P(du) := P_U(du) = \mathbb{P}\{U \in du\}$ and combine it with the likelihood $p(m|u) := p_{M|U}(m|u)$ using Bayes' rule to obtain the posterior $P(du|m) = \frac{p(m|u)P(du)}{p(m)}$, where $p(m) := \int p(m|u)P(du)$. If U had a density $p_U(u)$, then $P(du) = p_U(u)du$. Instead we use the general notation $P(du)$ for the distribution of U since a density exists for CIRF U only in a generalized sense. Specifically, at each site i there is positive mass for $u_i = 0$. Combined with our quadratic log-likelihood (5.8), the posterior is therefore

$$P_{(\epsilon)}(du) = Z_{(\epsilon)}^{-1} \exp(\epsilon \langle a + Au, u \rangle) P(du),$$

where $Z_{(\epsilon)} := \int \exp(\epsilon \langle a + Au, u \rangle) P(du)$ and the subscript with ϵ indicates the quadratic log-likelihood. Note that although we do not have an explicit expression for the prior $P(du)$ (and therefore no explicit expression for this posterior), we can indirectly use the prior for U through its moment generating functional (MGF). An even simpler way of using the MGF is via the cumulants. We start by applying the mean and covariance in §5.4 on linear filtering, and later use higher-order statistics for Volterra filters in §5.5. Finally, we exploit all orders of the CIRF statistics by explicitly using the MGF to represent a filter in §5.7.

5.4 The Linear Case

Although one may consider all statistical properties of the curve indicator random field, even the pair-wise relations express information critical to enhancing curves. As Gaussian random fields are fully defined by their means and covariances, in this section we simplify our contour filtering problem by replacing the CIRF prior with a Gaussian approximation $p_{\text{Gaussian}}(u)$ with the mean μ_U and covariance Σ_U from Corollary 4. Under this approximation, we can write the log-prior as

$$\begin{aligned} \ln p_{\text{Gaussian}}(u) &= -\frac{1}{2} \langle u - \mu_U, \Sigma_U^{-1} (u - \mu_U) \rangle + \text{constant}(m) \\ &= \langle \Sigma_U^{-1} \mu_U, u \rangle - \frac{1}{2} \langle \Sigma_U^{-1} u, u \rangle + \text{constant}(m). \end{aligned}$$

Combined with the quadratic log-likelihood (5.8) we get the following quadratic log-posterior (implying the posterior is Gaussian):

$$\begin{aligned} \ln p_{\text{Gaussian}}(u|m) &= \ln p(m|u) + \ln p_{\text{Gaussian}}(u) + \text{constant}(m) \\ &= \langle \epsilon a + \Sigma_U^{-1} \mu_U, u \rangle + \left\langle \left(\epsilon A - \frac{1}{2} \Sigma_U^{-1} \right) u, u \right\rangle + \text{constant}(m). \end{aligned}$$

Since the mean is equal to the mode for the Gaussian, we can obtain the posterior mean for the Gaussian approximation to the CIRF by differentiating the log-posterior with respect to u and then equating the result to 0. The posterior mean for the (general) quadratic log-likelihood is therefore

$$\widetilde{u_{\text{Gaussian}}} = (\Sigma_U^{-1} - 2\epsilon A)^{-1}(\epsilon a + \Sigma_U^{-1}\mu_U), \quad (5.9)$$

which is equal to $(\Sigma_U^{-1} + B^*\Sigma_N^{-1}B)^{-1}(B^*\Sigma_N^{-1}m + \Sigma_U^{-1}\mu_U)$ for the Gaussian log-likelihood [7] corresponding to the observation model $M = BU + N$, as in §5.2. The resulting posterior mean is a linear filtering of the measurements m by the operator $(\Sigma_U^{-1} + B^*\Sigma_N^{-1}B)^{-1}B^*\Sigma_N^{-1}$ (plus a fixed vector).

For the Gaussian likelihood case, this linear filter is known in image restoration as a Wiener filter [95]. The novelty in this application of the Wiener filter is that it operates in the higher-dimensional spaces where curves live, not two-dimensional intensity images. For the direction Markov process, this amounts to linear filtering in position (x, y) and orientation θ , and therefore our linear filter is an *oriented Wiener filter*. The domain of the random fields we are filtering is not the (discretized) image plane \mathbb{R}^2 , but instead the unit-length tangent bundle $\mathbb{R}^2 \times \mathbb{S}$. See Fig. 3.2 for an illustration of the kernel of the oriented Wiener filter for the case $B = I$ and $\Sigma_N = \sigma_N^2 I$ with $\sigma_N \rightarrow \infty$. When applied to the curvature Markov process, we obtain linear filtering in $\mathbb{R}^2 \times \mathbb{S} \times \mathbb{R}$. Other Markov process curve models will lead to filtering in other spaces.

Note again the presence of the measurement operator B . Using it, we can explicitly capture the imprecision in edges due to the finite aperture of edge operators. Clearly, we seek an estimate of the underlying curve indicator random field in which the edge strength is localized. In formal analogy to the Wiener filtering for restoring

out-of-focus grey-scale images, the filter here is a means of de-blurring edge measurement fields on $\mathbb{R}^2 \times \mathbb{S}$. Such a sharpening effect has been viewed as non-maximum suppression [39] or detection [51]. Other techniques for curve enhancement generally do not explicitly model blur at all. In contrast, blur is straightforward to include in this random field model.

Okajima [88] has independently proposed a closely-related Wiener filter for modeling the horizontal connections between complex cells in V1. Okajima’s filter operates in position and spatial frequency, and he exploits translation invariance to simplify his model, although he does not consider the additional rotation invariance suggested by $\mathbb{R}^2 \times \mathbb{S}$ and possessed by the oriented Wiener filter (see also §4.1). Interestingly, the coefficients of his model were learned from the responses of Gabor filters to a number of straight-line images and bear a qualitative resemblance to ours, which, in contrast, were derived from the blur operator B , the noise model, and curve indicator random field covariance.

5.4.1 Why Simple Cells and Local Edge Operators are Elongated

When Hubel and Wiesel reported the existence of cells in primate visual cortex that respond to edge and bar stimuli, they established a connection between physiology and geometry [45]. Unfortunately, no theory deriving the qualitative features of the receptive field for even their “simple cell” was provided. Recall that there are two key “directions” of the receptive field: normal and tangential to the major axis of the receptive field. Since the receptive fields of simple cells resemble the kernels of (oriented) local edge and line operators, understanding the rationale for one explains the other. The normal profile is generally understood functionally as a kind of blurred

differentiation [11, 93] using the theory of regularization. The tangential profile, with its elongated structure, is more poorly understood. Nonetheless many have designed elongated filters where the tangential profile is explained using a variety of theories, such as an uncertainty principle [16], Hilbert space representations [46], or increased noise immunity [11]. The CIRF linear filter corresponding to the direction process provides a much simpler explanation for the simple cell receptive field: it is tuned to find contour structure in the visual world. Using this idea, one can relate the length and aspect ratio of the simple cell receptive field to the expected length and curvature bound on contours that the cell can measure.

But does the curve indicator random field provide *more* than a contour-based explanation for elongated linear filtering? To answer this question, we shall have to go beyond the connection between linear filtering and second-order statistics.

5.5 Volterra Filters and Higher-Order Statistics

Volterra filters constitute one class of nonlinear filters that subsumes linear filters [107]. Even though most work on Volterra filtering is done on time-domain sequences, we shall apply these ideas to multidimensional signals (functions on $\mathbb{R}^2 \times \mathbb{S}$, for example) or on vectors $m = (m_i)$, where i is a site in \mathcal{I} . By definition, the output of a Volterra filter is a polynomial⁴ function of the input signal. Although Volterra filters are (generally) nonlinear, they can always be represented as a linear function of *products* of the input m . A k -th order Volterra filter is a linear function of 0- through k -fold products of the input; for example, a quadratic (second-order) Volterra filter can be written as a linear combination of a constant, the input m_i at site i , and the

⁴One must distinguish the nonlinear polynomial (Volterra) filters that we consider here from those linear filters whose *kernel* is based on some polynomial, e.g., a Hermite polynomial.

products $m_i m_j$ of input at sites i and j . A linear filter (plus a constant) is simply a first-order Volterra filter.

Analogous to the connection between MMSE linear filters and second-order statistics, MMSE Volterra filters are related to higher-order statistics. The key lies in exploiting the fact that Volterra filters are *linear* functions of products of the input. Instead of the actual inputs $\{m_i\}$, the idea is to focus on an “apparent” set of inputs $\{1\} \cup \{m_i\} \cup \{m_i \cdot m_j\} \cup \cdots \{m_{i_1} \cdot m_{i_2} \cdots m_{i_k}\}$, and so reduce the derivation of the (polynomial) Volterra filter on the actual inputs to the design of a linear filter on the apparent inputs. Because MMSE linear filters can be derived by solving a linear system (see §5.4) in terms of the second-order statistics of the input and the unknown U , one can therefore compute a MMSE Volterra filter by solving a linear system in terms of the second-order statistics of the apparent input and unknown, or the $(2k)$ -th order statistics of the input and unknown. In the context of discrete-time systems, such a linear system for Volterra filtering in terms of higher-order statistics was presented in [1] and later in [92, 10]. Based on empirical estimates of higher-order statistics, one-dimensional low-order (quadratic and occasionally cubic) Volterra filters have been designed [84, 134].

The generality of Volterra filtering can be immobilizing, for one must specify a large number of coefficients in designing one. In filtering a vector of length n (mathematically, a mapping from \mathbb{R}^n to \mathbb{R}^n), a linear filter requires $O(n^2)$ coefficients, whereas a general k -order Volterra filter requires $O(n^{k+1})$ coefficients. For applications in vision, the appropriate length n can be truly large: for images n is on the order of a million, and for the discretized $\mathbb{R}^2 \times \mathbb{S}$, n can be even a hundred times that. Not only is directly solving the linear system for the MMSE Volterra filter unthinkable in these dimensions, just estimating such higher-order statistics

requires enormous amounts of data. Applying the Volterra filter in the general case is similarly burdensome.

Symmetries and structural simplifications of any kind are essentially the only hope for taming the generality of the Volterra filter. One obvious symmetry is homogeneity: the kernel of the filter is the same irrespective of which portion of the image it is applied. This brings the number of coefficients need down to $O(n^k)$ from $O(n^{k+1})$ for the k -th order Volterra filter. Such homogeneity of the filter is a consequence of the statistical homogeneity of the unknown signal and noise, and amounts to stationarity for one-dimensional signals. Other simplifications arise when the unknown signal U is in some parametric family (e.g., ARMA models in one-dimension [84]) or when the Volterra filter is assumed to have a particular structure, such as factorizability [62] or a tensor product basis representation [87]. In vision, quadratic filters have been applied for local contour detection independent of phase (e.g., the contour's edge vs. line characteristics) [81], and general polynomial filters have been linked to the characterization and discrimination of textures [65]. Maass has made an interesting connection between Volterra filters and neural network dynamics [69]. In the case of contour filtering, we shall now see that the special structure of the CIRF cumulants (§3.4.3), combined with a Taylor series expansion in inverse noise level, makes MMSE Volterra filtering of noisy images possible.

5.6 Deriving CIRF Volterra Filters via Perturbation Expansion

Here we derive linear, quadratic and even cubic Volterra filters for the CIRF U observed in white Gaussian noise. In principle, similar techniques can also be used for filtering the CIRF observed with a log-likelihood polynomially dependent on U .

The basic idea will be to observe the dependence of the log-likelihood on a small parameter (the inverse noise variance), and to perform a perturbation expansion in that parameter. We shall study this perturbation expansion first, developing the notation and tools that will be necessary to carry it out.

Suppose we have a random vector $X = (X^i)$ (superscripts here index the component of the vector and do not represent taking a power) with prior distribution $P(dx)$ (with *known cumulants*) and posterior distribution $P_{(\epsilon)}(dx) = Z_{(\epsilon)}^{-1} \exp(\epsilon f(x)) P(dx)$, where ϵ is a small real number, $f(x) = \langle a + Ax, x \rangle$ is a real-valued function of x , $a = (a_i)$ is a constant vector, $A = (a_{ij})$ is a symmetric, negative definite linear operator, and the normalizing constant $Z_{(\epsilon)}$ is equal to $\int \exp(\epsilon f(x)) P(dx)$. Observe that this situation applies exactly to the CIRF in Gaussian white noise, where $U = X$, $\epsilon = \sigma_N^{-2}$, $a = m$, $a_{ij} = -\frac{1}{2}\delta_{ij}$ and $f(x) = f(u) = \langle m - \frac{1}{2}u, u \rangle$. Note that we do know all the cumulants of the CIRF prior U from Chapter 3. Given that that the main application for this thesis is the filtering of extremely noisy contour images, understanding how to optimally filter as $\sigma_N \rightarrow \infty$, or equivalently, as $\epsilon \rightarrow 0$, is very important. To avoid being distracted by the particulars of the application of this idea to the CIRF, let us now focus on the problem of estimating X . We can apply the results we derive to the CIRF later.

Recalling that the MMSE estimate of X is the posterior mean $\mathbb{E}_{(\epsilon)} X$, our task is to compute a Taylor series expansion of the posterior mean around $\epsilon = 0$. To foreshadow the results to come, we shall see that up to order ϵ^3 the expansion of the posterior mean to order ϵ^k is a k -th order Volterra filter requiring cumulants of up to the $(k + 1)$ -th order for the CIRF.⁵

⁵This assumes that the CIRF is based on an approximately self-avoiding Markov process; otherwise, $(2k + 1)$ -th order cumulants would be required.

Now, observe that the (posterior) mean can be computed by differentiating the logarithm of the normalizing constant, or

$$\frac{\partial \ln Z_{(\epsilon)}}{\partial(\epsilon a_i)} = Z_{(\epsilon)}^{-1} \int \frac{\partial \exp(\epsilon \langle a + Ax, x \rangle)}{\partial(\epsilon a_i)} P(\mathrm{d}x) = \int x_i P_{(\epsilon)}(\mathrm{d}x) = \mathbb{E}_{(\epsilon)} X^i. \quad (5.10)$$

Therefore to get a Taylor series for the posterior mean, we need only get a Taylor series for $\ln Z_{(\epsilon)}$ and then differentiate with respect to ϵa_i .

Where do we get a Taylor series for $\ln Z_{(\epsilon)}$? Observe that the normalizing constant $Z_{(\epsilon)}$ is just the moment generating functional of the random vector $Y = f(X)$ in disguise, or, equivalently, Y has the cumulant generating functional $\text{CGF}_Y(\epsilon)$ such that

$$\ln Z_{(\epsilon)} = \ln \int \exp(\epsilon f(x)) P(\mathrm{d}x) = \ln \int \exp(\epsilon y) P(\mathrm{d}x) = \ln \text{MGF}_Y(\epsilon) = \text{CGF}_Y(\epsilon). \quad (5.11)$$

But the coefficients in a Taylor series expansion of a cumulant generating functional are by definition the cumulants, and therefore we need to obtain the cumulants of $Y = f(X)$, a (quadratic) polynomial in X . Unfortunately, the cumulants of Y are not directly known; we must somehow calculate them in terms of the (known) cumulants of X .

5.6.1 Tensors and Generalized Cumulants

Without appropriate tools, calculating the cumulants of a polynomial can be extraordinarily difficult. To aid in this task, McCullagh has developed a tensor formalism for representing moments, cumulants, and so-called generalized cumulants [75, 76]. Elsewhere in early vision, tensors have previously been used to represent blurred image derivatives via the multiscale local jet [33, 31, 32] and to summarize local tangent

information for relaxation labeling [21] and linear filtering (“tensor voting”) [39]. We now review McCullagh’s formalism.

For the purposes of this discussion, a tensor is an indexed array of numbers, for example a_{ij} or b_k^{ij} , where indices such as i, j, k can appear as superscripts or subscripts and vary over a set of index values, e.g., $i \in \mathcal{I}$. Tensors are multiplied together by multiplying the corresponding array entries (as numbers), as in $a_i X^i$ or $b_{ij} d_i$, although we shall generally adopt the summation convention. In this convention, an index that appears in both a subscript and a superscript is assumed to be summed over; e.g., $a_i X^i := \sum_i a_i X^i$, and $a_{ij} Y^j := \sum_j a_{ij} Y^j$. An index that appears twice or more as a superscript, but not as a subscript (or vice versa) is not assumed to be summed over; so, for example, we do *not* assume that $g^{ij} g^{jk}$ is equal to $\sum_j g^{ij} g^{jk}$, but we do assume that $a_j g^{ij} g^{jk} := \sum_j a_j g^{ij} g^{jk}$. The summation convention often makes for pithy expressions, but we shall not hesitate from writing out the sum explicitly if it helps to reduce ambiguity.

Moments and cumulants can conveniently be represented using tensors, as follows. The first moment $\mathbb{E}X^i$ will be denoted by κ_X^i and the second moment $\mathbb{E}X^i X^j$ by κ_X^{ij} . More generally, the k -th moment $\mathbb{E}X^{i_1} \cdots X^{i_k}$ of the i_1 -th through i_k -th components of X will be indicated by $\kappa_X^{i_1 \cdots i_k} = \kappa^{i_1 \cdots i_k}$, where the subscript X will be dropped if understood. Cumulants will be written in a similar fashion, except that commas will separate the indices; we define $\kappa^{i,j,k} := \text{cum}\{X^i, X^j, X^k\}$, for example (where X is understood). In general, the k -th cumulant of X^{i_1} through X^{i_k} is denoted by $\kappa^{i_1, \dots, i_k} := \text{cum}\{X^{i_1}, \dots, X^{i_k}\}$. Observe that κ^i denotes both the first moment and first cumulant of X^i , which is appropriate since they are equal.

A *generalized cumulant* of random vector X is simply a cumulant of products of the components of X , such as $\text{cum}\{X^i, X^j X^k\}$. McCullagh has suggested the

convenient notation

$$\kappa^{ij,k,lmn} := \text{cum}\{X^i X^j, X^k, X^l X^m X^n\} \quad (5.12)$$

for generalized cumulants, where the commas in the superscript separate blocks of indices, with each block corresponding to a product of components of X and where these products are separated by commas in the (ordinary) cumulant (as on the right hand side of (5.12)). Due to the symmetry of both (ordinary) cumulants and products of components in their indices, the generalized cumulant is symmetric in its arguments in two senses: (1) the order of the blocks is irrelevant and (2) the order of the indices within each block is irrelevant. Indices cannot be moved from one block to another, however, and therefore $\kappa^{i,jk}$ is equal to $\kappa^{kj,i}$ but not equal to $\kappa^{ij,k}$, for example. Observe that an ordinary cumulant of k -th order is a generalized cumulant with k blocks of indices, with just one index per block, or κ^{i_1, \dots, i_k} . In contrast, a k -th order moment is simply a generalized cumulant with one block of k indices, or $\kappa^{i_1 \dots i_k}$. Thus the generalized cumulant indeed is a generalization of the statistical constructions of moments and cumulants.

How can a generalized cumulant be expressed in terms of ordinary cumulants? McCullagh has derived a formula for generalized cumulants as a sum of products of ordinary cumulants. Because we use this formula later, we shall set it up here. Observe that a generalized cumulant with α blocks involving a total of β random variables can be written as $\kappa(\mathfrak{P})$, where $\mathfrak{P} := \{v_1, \dots, v_\alpha\}$ is a partition of β indices into α non-empty blocks v_1, \dots, v_α of indices. Now, any such partition can be written as a graph with β vertices, where one vertex is for each random variable or random vector component. All vertices corresponding to each block v_j of the partition will be connected by edges, but no two vertices in distinct blocks will be linked by an

edge. In this way, a partition \mathfrak{P} corresponds to a graph that is a union of α complete graphs, but with no two such complete subgraphs being connected. Thus \mathfrak{P} will also denote this corresponding graph. Given any other partition \mathfrak{P}' of the same set of indices (and its corresponding graph), the edge sum $\mathfrak{P} \oplus \mathfrak{P}'$ is a graph in the same vertices where the set of edges is the union of the edges for \mathfrak{P} and \mathfrak{P}' . Now we can state McCullagh's formula for the generalized cumulant in terms of products of ordinary cumulants:

$$\kappa(\mathfrak{P}) = \sum_{\mathfrak{P}'} \kappa(v'_1) \cdots \kappa(v'_n), \quad (5.13)$$

where the sum extends over all n -block partitions $\mathfrak{P}' := \{v'_1, \dots, v'_n\}$ such that $\mathfrak{P} \oplus \mathfrak{P}'$ is a connected graph. The factors $\kappa(v'_j)$ in the sum are ordinary cumulants in the indices in block v'_j of the partition \mathfrak{P}' .

To illustrate McCullagh's result, consider the generalized cumulant corresponding to the partition $\mathfrak{P} = \{\{1, 2\}, \{3, 4\}, \{5\}, \{6\}\}$ with $\alpha = 4$ blocks and $\beta = 6$ indices. While each block of partition $\mathfrak{P}' = \{\{1, 3\}, \{2, 4\}, \{5, 6\}\}$ joins a block of \mathfrak{P} , the edge sum $\mathfrak{P} \oplus \mathfrak{P}'$ is not connected (because $\{5, 6\}$ is disconnected from the other nodes) and so there is no term in the sum in (5.13) corresponding to \mathfrak{P}' . However, the partition $\mathfrak{P}'' = \{\{1, 2, 3\}, \{4, 5, 6\}\}$ does contribute the term $\kappa(\{1, 2, 3\})\kappa(\{4, 5, 6\}) = \kappa^{1,2,3}\kappa^{4,5,6}$ to the sum because $\mathfrak{P} \oplus \mathfrak{P}''$ is connected. Using the formula, one obtains, for example,

$$\kappa^{i,jk} = \kappa^{i,j,k} + \kappa^{i,j}\kappa^k + \kappa^{i,k}\kappa^j = \kappa^{i,j,k} + \kappa^{i,j}\kappa^k[2],$$

where the bracketed number on the right of a term is the number of partitions of similar form that should be included in the sum. In this case, $\kappa^{i,j}\kappa^k[2]$ equals $\kappa^{i,j}\kappa^k + \kappa^{i,k}\kappa^j$. Although this bracket notation can be ambiguous (e.g., note that term $\kappa^{j,k}\kappa^i$, although of similar form, is not implied by the above use of $\kappa^{i,j}\kappa^k[2]$), it is useful in avoiding truly long enumerated lists. Another example of McCullagh's

formula is

$$\kappa^{i,j,kl} = \kappa^{i,j,k,l} + \kappa^k \kappa^{i,j,l}[2] + \kappa^{i,k} \kappa^{j,l}[2].$$

Applied to the third moment κ^{ijk} , the formula requires a sum over all partitions of $\{i, j, k\}$ (because $\mathfrak{P} = \{\{i, j, k\}\}$ is already connected); and applied to ordinary cumulant $\kappa^{i,j,k,l}$ (with $\mathfrak{P} = \{\{1\}, \{2\}, \{3\}, \{4\}\}$), the formula is satisfied by exactly one partition $\mathfrak{P}' = \{\{1, 2, 3, 4\}\}$ connecting all nodes. Thus McCullagh's formula succinctly expresses generalized cumulants in terms of a sum of products of ordinary cumulants.

5.6.2 Cumulants of Polynomials

With the generalized cumulant formula, we now need only express the cumulants of our polynomial $Y = f(X)$ in terms of the generalized cumulants of X . By the linearity of expectation and of cumulants [112], cumulants of homogeneous polynomials such as $a_i X^i$ and $a_{ij} X^i X^j$ can be computed straightforwardly. For example, $\mathbb{E}[a_i X^i] = \text{cum}\{a_i X^i\} = a_i \text{cum}\{X^i\} = a_i \kappa^i$ and

$$\begin{aligned} \text{var}(a_{ij} X^i X^j) &= \text{cum}\{a_{ij} X^i X^j, a_{kl} X^k X^l\} = a_{ij} a_{kl} \text{cum}\{X^i X^j, X^k X^l\} = a_{ij} a_{kl} \kappa^{ij,kl} \\ &= a_{ij} a_{kl} (\kappa^{i,j,k,l} + \kappa^{i,k} \kappa^{j,l}[2] + \kappa^{i,j} \kappa^{k,l}[2] + \kappa^k \kappa^{i,j,l}[2] + \kappa^i \kappa^k \kappa^{j,l}[4]), \end{aligned}$$

using the generalized cumulant formula. Although one can compute the cumulants of general polynomial transformations of a random variable [76, 62 ff.], here we focus on our case of interest, $Y = \langle a + AX, X \rangle = a_i X^i + a_{ij} X^i X^j =: (A_1 + A_2)X =: \mathcal{P}X$, where we use the notation $\mathcal{P} = A_1 + A_2$ for our polynomial operator with linear part A_1 and quadratic part A_2 . One can [76, p. 63] express the cumulant generating

functional in terms of the cumulant tensors κ of X using the formal expansion

$$\text{CGF}_Y(\epsilon) = \exp(\epsilon \mathcal{P})\kappa = (1 + \epsilon \mathcal{P} + \epsilon^2 \mathcal{P}\mathcal{P}/2! + \epsilon^3 \mathcal{P}\mathcal{P}\mathcal{P}/3! + \dots)\kappa.$$

The terms in this expansion are determined by expanding the polynomial operator into a sum of homogeneous polynomials, as follows:

$$\mathcal{P}\kappa = (A_1 + A_2)\kappa = a_i \kappa^i + a_{ij} \kappa^{ij}$$

and

$$\begin{aligned} \mathcal{P}\mathcal{P}\kappa &= (A_1 + A_2)(A_1 + A_2)\kappa \\ &= A_1 A_1 \kappa + A_1 A_2 \kappa + A_2 A_1 \kappa + A_2 A_2 \kappa \\ &= a_i a_j \kappa^{i,j} + a_i a_{jk} \kappa^{i,jk} + a_{ij} a_k \kappa^{ij,k} + a_{ij} a_{jk} \kappa^{ij,jk} \\ &= a_i a_j \kappa^{i,j} + a_i a_{jk} \kappa^{i,jk}[2] + a_{ij} a_{jk} \kappa^{ij,jk}. \end{aligned}$$

Observe there is only one block of indices in the generalized cumulants in $\mathcal{P}\kappa$, but there are two blocks for the generalized cumulants in $\mathcal{P}\mathcal{P}\kappa$. Higher order terms behave similarly, as we now observe in our sought-for expression of the cumulant generating function of the polynomial Y in terms of the generalized cumulants of X :

$$\begin{aligned} \text{CGF}_Y(\epsilon) &= \epsilon (a_i \kappa^i + a_{ij} \kappa^{ij}) + \frac{\epsilon^2}{2} (a_i a_j \kappa^{i,j} + a_i a_{jk} \kappa^{i,jk}[2] + a_{ij} a_{kl} \kappa^{ij,kl}) \\ &\quad + \frac{\epsilon^3}{6} (a_i a_j a_k \kappa^{i,j,k} + a_i a_j a_{kl} \kappa^{i,j,kl}[3] + a_i a_{jk} a_{lm} \kappa^{i,jk,lm}[3] + a_{ij} a_{kl} a_{mn} \kappa^{ij,kl,mn}) \\ &\quad + \frac{\epsilon^4}{24} (a_i a_j a_k a_l \kappa^{i,j,k,l} + a_i a_j a_k a_{lm} \kappa^{i,j,k,lm}[4] + a_i a_j a_{kl} a_{mn} \kappa^{i,j,kl,mn}[6] \\ &\quad + a_i a_{jk} a_{lm} a_{no} \kappa^{i,jk,lm,no}[4] + a_{ij} a_{kl} a_{mn} a_{op} \kappa^{ij,kl,mn,op}) + O(\epsilon^5). \end{aligned}$$

This expression can be further simplified by observing that all of the bracketed numbers representing permutations become scalar factors by relabeling indices appropriately. To obtain the posterior mean $\mathbb{E}_{(\epsilon)}X$, by (5.10) and (5.11) we need only differentiate $\text{CGF}_Y(\epsilon)$ with respect to ϵa_r , as follows:

$$\begin{aligned}
\mathbb{E}_{(\epsilon)}X^r &= \kappa^r \\
&+ \epsilon \left(a_i \kappa^{r,i} + a_{ij} \kappa^{r,ij} \right) \\
&+ \frac{\epsilon^2}{2} \left(a_i a_j \kappa^{r,i,j} + 2a_i a_{jk} \kappa^{r,i,jk} + a_{ij} a_{kl} \kappa^{r,ij,kl} \right) \\
&+ \frac{\epsilon^3}{6} \left(a_i a_j a_k \kappa^{r,i,j,k} + 3a_i a_j a_{kl} \kappa^{r,i,j,kl} + 3a_i a_{jk} a_{lm} \kappa^{r,i,jk,lm} \right. \\
&\quad \left. + a_{ij} a_{kl} a_{mn} \kappa^{r,ij,kl,mn} \right) \\
&+ O(\epsilon^4).
\end{aligned} \tag{5.14}$$

Expression (5.14) can be used directly to compute the posterior mean of a signal (with known cumulants) that is corrupted by linear blur and (the limit of) large, possibly correlated, Gaussian noise (for more general observation models, see §5.2.2). One can derive linear ($O(\epsilon)$ terms), quadratic ($O(\epsilon^2)$ terms), cubic ($O(\epsilon^3)$ terms) and other higher-order Volterra filters by systematically computing all of the generalized cumulants of X in terms of its ordinary cumulants.

5.6.3 The CIRF: A Tractable Special Case

In the general case, the expansion of formula (5.14) is a lengthy and tedious task: each generalized cumulant will become a possibly large number of products of ordinary cumulants. In addition, the resulting computation naively suggested by the summation convention would have high-order polynomial complexity, where the order depends on the number of terms in the perturbation expansion, e.g., $O(n^7)$ operations for the

cubic filter, which has all the terms explicitly listed in (5.14). For $n = 10^6$, applying the cubic filter would naively require on the order of $(10^6)^7 = 10^{42}$ operations. It is little wonder, therefore, that this formula has not appeared in high-noise signal processing or vision applications before.

Fortunately, the contour images modeled by the CIRF have a special cumulant structure that we can exploit (see §3.4.3). In addition, we invoke several realistic assumptions that will both (1) significantly simplify the task of even writing down the surfeit of terms and (2) reduce the computational complexity of the filters to that of applying the Green's operator G . For the direction process-based CIRF, the filter complexity becomes $O(n \log n)$, where $n = |\mathcal{I}|$. We now consider these extra assumptions in turn.

Spatial Homogeneity

For general purpose contour enhancement, we assume that any knowledge about contour endpoint locations will be conveyed by the contour image itself. Equivalently, one is indifferent to the starting and ending locations of each contour. As introduced in §3.4.2, this amounts to setting the initial distribution μ and final weighting ν proportional to a constant, e.g., $\mu_i = |\mathcal{I}|^{-1}$, $\nu_i = \lambda^{-1}$, for all sites $i \in \mathcal{I}$. Using the notion of contour *density* $\eta := \bar{\mathcal{N}}\lambda/|\mathcal{I}|$ (the average amount of contour per site) and observing that $G\mathbf{1} = \lambda\mathbf{1}$, a simple calculation shows that the k -th cumulant of the spatially homogeneous CIRF U at sites i_1, \dots, i_k becomes:

$$\kappa^{i_1, \dots, i_k} = \text{cum}\{U_{i_1}, \dots, U_{i_k}\} = \eta \sum_{a \in \text{perm}_k} g^{i_{a_1} i_{a_2}} \dots g^{i_{a_{k-1}} i_{a_k}}. \quad (5.15)$$

(Superscripts are used here in keeping with the tensor notation and summation convention.) In addition, the Markov process R_t itself will be assumed to be spatially

homogeneous in the sense that shifts⁶ in the initial conditions give rise to corresponding shifts in the solution to the corresponding Fokker-Planck diffusion. As a consequence, the Green's operator $G = (g^{ij})$ is constant along the diagonal, i.e., $g^{ii} =: d$, for all $i \in \mathcal{I}$.

Low Contour Density

Another fact that we shall exploit is that, in a large number of images, contours constitute a small fraction of the total image. For the CIRF, this amounts to asserting that the contour density η is small ($0 \leq \eta \ll 1$). In the continuum, an extreme form of this claim certainly holds: the bounding contours in an image of smooth surfaces in an occlusion relationship constitute a set of measure zero (with respect to the standard area measure in the plane). In natural images it is easy to observe a variety of sources of contours for which the low-density CIRF is appropriate, such as occlusion edges, highlights, surface markings (albedo discontinuities), and shadow boundaries. A number of textures, in particular those where contour structure dominates regions, may not appear to be modeled realistically by low-density CIRFs, and indeed this is the case if the underlying Markov process is planar Brownian motion. But when using higher-dimensional Markov process contour models, even images with a large number of crossing contours can be reasonably modeled with low-density CIRFs. For example, contours that cross in a two-dimensional image will not cross in $\mathbb{R}^2 \times \mathbb{S}$ unless they have the same orientation at the crossing point. Working with curvature in $\mathbb{R}^2 \times \mathbb{S} \times \mathbb{R}$, even highly dense *planar* texture patterns are represented with lower density as well.

⁶Depending on the Markov process, the shifts might include translations and rotations (e.g., planar Brownian motion and Mumford's direction process). In general the shifts will be the set of transformations under which the Fokker-Planck equation is invariant (see [140] for some discussion of this).

The key consequence of low contour density is that the generalized cumulants in (5.14) can be approximated with one corresponding joint cumulant, as follows. Recall formula (5.13) for expanding generalized cumulant $\kappa(\mathfrak{P})$, where graph $\mathfrak{P} = \{v_1, \dots, v_\alpha\}$ has vertices i_1, \dots, i_k , into a sum of products of ordinary cumulants corresponding to partitions \mathfrak{P}' such that $\mathfrak{P} \oplus \mathfrak{P}'$ is connected. Observe that the complete graph $\mathfrak{P}'' = \{\{i_1, \dots, i_k\}\}$ (with only one block in the corresponding partition) always leaves $\mathfrak{P} \oplus \mathfrak{P}''$ connected, since \mathfrak{P}'' is connected. Thus there will always be the term $\kappa(\mathfrak{P}'') = \kappa^{i_1, \dots, i_k}$ (a *single* ordinary cumulant in all the vertices of \mathfrak{P}) in formula (5.13), and by (5.15) this term will be $O(\eta)$. All other terms in the generalized cumulant expansion will be products of ordinary cumulants, and therefore will be $O(\eta^2)$. As a result, with an error of $O(\eta^2)$ we shall replace all of the generalized cumulants in (5.14) with the corresponding joint ordinary cumulant. For example, $\kappa^{r,ij} = \kappa^{r,i,j} + O(\eta^2)$ and $\kappa^{r,i,jk,lm} = \kappa^{r,i,j,k,l,m} + O(\eta^2)$ for the low-density CIRF. When the low-density assumption does not hold, the generalized cumulant expansion can be evaluated completely. Alternatively, one could even consider a high contour density limit where only those terms of highest power in the contour density η are kept.

Self-Avoidingness of Individual Contours

Suppose our Markov processes were perfectly straight. Then these “curves” would never return to their starting points even if they were infinitely long. Smooth contours approximately obey this relationship for short lengths, and our Markov process contour models have similar behaviour. The direction process, for example, rarely returns to its starting position; the curvature process only returns after a full, smooth, loop, unlike planar Brownian motion, which often self-intersects immediately after passing through a site. This “approximately self-avoiding” nature of our Markov

process models will be exploited in deriving the Volterra filters, allowing many terms to be neglected, as we shall see now.

5.6.4 Diagram Method for Evaluating the Posterior Mean

Let us now evaluate the terms in parentheses in the posterior mean expansion (5.14). For the case of additive Gaussian white noise, recall that $a_{ij} = -\frac{1}{2}\delta_{ij}$. After replacing the generalized cumulants with ordinary cumulants via the low contour density assumption, our task is to compute terms such as $a_i a_j \kappa^{r,i,j}$ and $a_i \delta_{jk} \delta_{lm} \kappa^{r,i,j,k,l,m}$. Also, to avoid unnecessary distraction, we suppress all multiplicative powers of $-\frac{1}{2}$ from the a_{ij} s and the scalar multiplication by η in the CIRF form of the cumulants⁷ until we add up the evaluated terms later. Now, evaluating these terms requires expanding the (ordinary) cumulants $\kappa^{r,i,j,\dots}$ into a sum over all permutations of the indices by formula (5.15). The last term for the cubic filter, for example, will require that $7! = 5040$ permutations be evaluated.

To support the error-prone task of explicitly summing over each permutation and to aid the intuition, we shall construct a diagram corresponding to each permutation. In statistical physics and quantum field theory, Feynman diagrams have been successfully used as a computational device to support perturbation calculations. Independently of this work, Ruderman [103, pp. 29–33] used Feynman diagrams to calculate the posterior mean of an underlying signal observed in Gaussian white noise, where the perturbation arises from the non-Gaussianity of the underlying signal. In contrast, the perturbation used here is due to the Gaussian noise; the underlying non-Gaussian signal (the CIRF) is modeled exactly. In addition, the diagram we construct is not a Feynman diagram.

⁷Set $\eta = 1$ in (5.15) until the final sum of terms in (5.14).

Let us start with the simple case of $a_i \kappa^{r,i} = a_i g^{ri} + a_i g^{ir}$, using (5.15). This sum has two terms, the first corresponding to the permutation (r, i) . For each term we construct our diagram as a graph with one node per index, and with directed edges corresponding to the order of the permutation (Fig. 5.3 (a)). Observe that no node occurs twice in the permutation, and so the diagram just represents an ordered, simple path along the nodes. We then embellish the diagram (Fig. 5.3 (b)) to illustrate how it corresponds to the expression $a_i g^{ri}$: node r (the “readout”) is identified with a hollow dot (“observation”), and the nodes corresponding to the coefficient a_i are marked with an “X” to denote an “amplification,” “boost,” or “interaction” with input a .

The term $a_i g^{ri}$ corresponding to this diagram (Fig. 5.3 (a)) represents the action of operator G on vector a and evaluated at r , or $(Ga)^r$; observe that the edge is directed *away from* r . Conversely, the term $a_i g^{ir} = (G^*)^{ri} a_i$ corresponding to permutation (i, r) represents the action of operator G^* on vector a and evaluated at r , with the edge directed *towards* r . One can check that this pattern persists for more complex diagrams and so we can write the operator G along each link directed away from r and G^* along each link directed towards r . Dropping the r to represent the vectorial result, we work inwards toward the readout r on the diagram to simply write Ga for permutation (r, i) , and G^*a for (i, r) .

For term $a_i a_j \kappa^{r,i,j}$, we need to consider three kinds of permutations of the indices r, i , and j . Permutation (r, i, j) corresponds to $a_i a_j g^{ri} g^{ij}$, which can be evaluated as $g^{ri} a_i (Ga)^j = g^{ri} (\text{diag } a Ga)^i = G \text{diag } a Ga$ because the componentwise product $a_i b_i$ of any vectors a and b can always be written as $(\text{diag } a b)_i$, the i -th component of the matrix $\text{diag } a$ times vector b . This result can be read directly from Fig. 5.3 (d) by working from node j towards the readout r : the input a comes in first at j , is operated on by G , the result Ga is componentwise multiplied by a (equivalently, operated upon

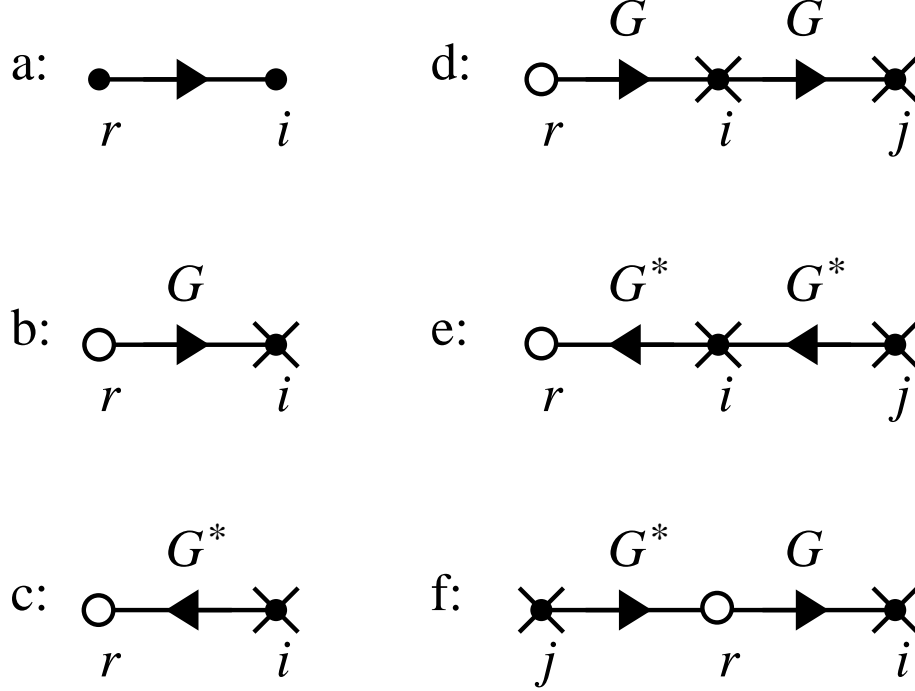


Figure 5.3: Diagrams for representing terms of the posterior mean (5.14). Each diagram corresponds to a particular permutation (via (5.15)) for one of the terms. The permutation for each diagram is a: (r, i) ; b: (r, i) ; c: (i, r) ; d: (r, i, j) ; e: (j, i, r) , and f: (j, r, i) . The nodes represent indices in the term, and the edges represent the ordering of indices in the corresponding permutation from the CIRF cumulant. The hollow nodes represent the output node (not summed over); the “X” indicates the application of an input a at the node.

by diag a), which is finally operated upon by G , to produce $G \text{ diag } a Ga$. Since we get the same diagram back when nodes i and j are swapped, to permutation (r, j, i) we get the same result $G \text{ diag } a Ga$. Permutation (j, i, r) has the diagram shown Fig. 5.3 (e), which is the same as the reversal of the diagram for permutation (r, i, j) , and so corresponds to $G^* \text{ diag } a G^* a$. Again, nodes i and j can be swapped to produce a $G^* \text{ diag } a G^* a$ for permutation (i, j, r) . The readout r need not be at the end of the ordering, as in permutation (j, r, i) corresponding to term $a_i a_j g^{jr} g^{ri} = (G^* a)^r (Ga)^r$. Letting $a \odot b$ denote the vector which is the componentwise product of vectors a and b , i.e., $(a \odot b)_i = a_i b_i$, the vector result from permutation (j, r, i) can therefore be written as $G^* a \odot Ga := (G^* a) \odot (Ga)$. This can also be derived from the diagram

by splitting the diagram (Fig. 5.3 (f)) at r , separately treating the two resulting sub-paths (one directed into r , the other directed away from r) as before, and then taking the componentwise product as the result. Note that i and j can be swapped as before for another $G^*a \odot Ga$. Gathering together the results from Fig. 5.3 (d-f), we obtain

$$a_i a_j \kappa^{r,i,j} = (2G \text{ diag } a \text{ } Ga + 2G^* \text{ diag } a \text{ } G^*a + 2G^*a \odot Ga)^r.$$

In hindsight, this result can be derived most easily from the appropriate diagrams.

In addition to terms with only coefficients of form a_i , we also have to deal with terms that have a Kronecker delta, such as $a_i \delta_{jk} \kappa^{r,i,j,k}$. Because the Kronecker delta δ_{jk} evaluates to 1 only when the two indices have equal value, and zero otherwise, the coefficient δ_{jk} represents the identity of indices j and k . In the diagram this amounts to the collapsing of nodes j and k into a new combined node $\{j, k\}$, giving rise to a loop. For permutation (r, i, j, k) , we therefore get the diagram (a) in Fig. 5.4 for the expression $a_i \delta_{jk} g^{ri} g^{ij} g^{jk} = g^{ri} a_i \sum_j g^{ij} g^{jj}$. To evaluate this diagram, we start at node $\{j, k\}$. We first realize that, due to spatial homogeneity, the G is constant along the diagonal, or $g^{jj} = d = (d\mathbf{1})_j$ for all j . On the diagram this can be represented by replacing the loop with an “input” of $d\mathbf{1}$ at node $\{j, k\}$ (all one-edge loops can be replaced in this way). Working backwards towards r , we find that permutation (r, i, j, k) evaluates to $G \text{ diag } a \text{ } Gd\mathbf{1} = d\lambda Ga$, since $G\mathbf{1} = \lambda\mathbf{1}$. Having evaluated this diagram we know how to evaluate similar diagrams. First, the reversal of nodes j and k produces $d\lambda Ga$ for permutation (r, i, k, j) . Second, reversing all of the arrows (for both orderings of j and k) causes the substitution of operator G^* for G .

For permutation (j, k, r, i) we must evaluate $a_i \delta_{jk} g^{jk} g^{kr} g^{ri}$. Explicitly summing over indices we obtain $(Ga)^r \sum_j g^{rj} g^{jj} = d\lambda(Ga)^r$. But from the diagram (Fig. 5.4 (b)),

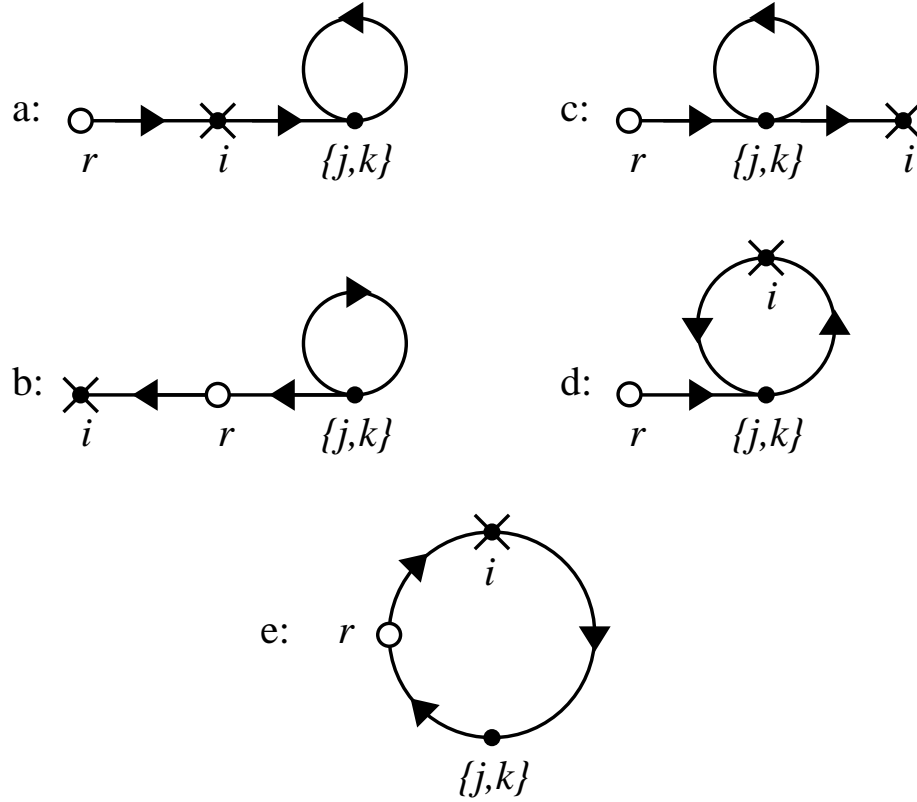


Figure 5.4: Diagrams with loops for representing permutations in the CIRF cumulant formula (5.15) for expression $a_i \delta_{jk} \kappa^{r,i,j,k}$ from the posterior mean (5.14). Terms corresponding to diagrams (a: permutation (r, i, j, k)) and (b: (j, k, r, i)) will be kept, whereas we shall neglect terms corresponding to diagrams (c: (r, j, k, i)) (by the input-output-coupling rule), (d: (r, j, i, k)) and (e: (j, r, i, k)) (by the 1-link-loop rule). Only these two rules are used to decide whether to keep or drop a particular permutation.

we can just take the componentwise product of the results due to the path towards and the path away from r , with input a at node i and input $d\mathbf{1}$ (for the loop) at node $\{j, k\}$. We then get the vectorial result $Ga \odot dG\mathbf{1} = d\lambda Ga$. Again, nodes j and k can be swapped and the entire path can be reversed (G becomes G^*).

For permutation (r, j, k, i) , we need to evaluate $a_i \delta_{jk} g^{rj} g^{jk} g^{ki}$. Again, the loop becomes an input $d\mathbf{1}$ at $\{j, k\}$, and so we can use the diagram (Fig. 5.4 (c)) to obtain the result $G \text{ diag } d\mathbf{1} Ga = dGIGa = dGGa$.

Now, observe that G is a positive operator for the Markov processes in which we are interested⁸. Given that there is only one positive eigenvector for a positive operator and that it corresponds to the maximum eigenvalue [55, p. 67], the equation $G\mathbf{1} = \lambda\mathbf{1}$ implies that λ is the maximum eigenvalue because $\mathbf{1}$ is a positive eigenvector of G . This implies that $\lambda^{-1}G$ is a contraction mapping that only passes constant vectors unchanged; non-constant vectors will shrink. Equivalently, $\|Gy\| < \|\lambda y\|$ unless y is a constant vector.

Because of this, we view the result $dGGa = dGy$ for permutation (r, j, k, i) as smaller than the result $d\lambda Ga = d\lambda y$ for permutation (j, k, r, i) , where $y = Ga$ is generally non-constant for typical inputs a . Observe that such contraction mappings will correspond to repeated operators (GG , G^*G^* , etc., without a “boost” or “amplification” from a between them) whenever there is a loop located *between* a regular input a (located at i for permutation (r, j, k, i)) and the output r . As a first approximation, we shall therefore neglect $dGGa$ and all other permutations that have a Kronecker delta index between an input node and the output r ; we shall restrict our attention to those permutations where all of the indices with subscripted inputs a form an uninterrupted sequence (in the permutation) to the output r . Let us call this criterion for neglecting terms the *input-output-coupling rule*.

In addition to this rule for neglecting loops, there is another that is due to the approximately self-avoiding nature of our Markov processes. We illustrate this rule with permutation (r, j, i, k) (Fig. 5.4(d)), or term $a_i \delta_{jk} g^{rj} g^{ji} g^{ik} = \sum_j g^{rj} (G^{ji} (G^*)^{ji}) a_i = (G(G \odot G^*)a)^r$, where $A \odot B$ denotes the componentwise product of the two operators A and B . Here, the effective input at combined node $\{j, k\}$ due to the 2-edge loop $\{\{j, k\}, i\}$ is $(G \odot G^*)a$. We have “broken” the loop at i and taken the componentwise

⁸Equivalently, the Markov process on average spends nonzero time in every state j given that it started in state i , for both the direction and curvature processes; this can be seen from the plots of the Green’s operators, which—although they become very small—do not go exactly to zero away from the impulse location (Figs. 2.2 and 2.4).

product of the operators that correspond to the two distinct paths from i to $\{j, k\}$; this is the action of operator $G \odot G^*$ on vector a . For approximately self-avoiding processes, observe that $G \odot G^*$ is approximately $d^2 I$: each row is the componentwise product of the corresponding rows of G and G^* , which, by spatial homogeneity, is just the componentwise product of the Green's operator kernels for the forward operator G and the backward operator G^* . For the directed Markov process in $\mathbb{R}^2 \times \mathbb{S}$, the forward operator is large only on one side of the impulse, and the backward operator is large only on the other, and thus the componentwise product is only significant at the origin, which corresponds to the common diagonal $g^{ii} = d$. For the curvature process in $\mathbb{R}^2 \times \mathbb{S} \times \mathbb{R}$, the forward and backward kernels may overlap for impulses of high curvature, suggesting the need for a bound on curvature. Thus permutation (r, j, i, k) evaluates to the simple expression $d^2 Ga$, which is clearly smaller than the $d\lambda Ga$ expression for permutation (r, i, j, k) , since $d = \tau$ (the average amount of time the contour spends in a site as it passes through) is certainly much less than the total average contour length λ . This example suggests the *1-link-loop rule*: neglect those permutations that correspond to diagrams where the largest loop has more than one link or edge.

5.6.5 Computational Significance of Approximately Self-Avoiding Processes

It is interesting to note that for Brownian motion (a symmetric process, with $G^* = G$) there is perfect overlap between the forward and backward kernels, so the above argument would not apply and therefore we should not invoke the 1-link-loop rule to neglect such terms with inputs inside a loop. This suggests a novel justification for orientation in vision: using approximately self-avoiding Markov processes (e.g.

the directed process in $\mathbb{R}^2 \times \mathbb{S}$) gives rise to a reduction in computational complexity compared to non-self-avoiding processes such as planar Brownian motion. Evolution might exploit this fact since those organisms that expend fewer resources on contour computations are more likely to survive. Observe that this is rather counter-intuitive because it is the more elaborate representation $(\mathbb{R}^2 \times \mathbb{S})$, not the simpler (\mathbb{R}^2) , that has the smaller computational burden. Also note that this complexity argument is distinct from the statistical argument as to whether contours in the world do tend to smoothly continue (§4.2) in orientation.

Returning to the calculation, both the 1-link-loop rule and the input-output-coupling rule force us to drop permutation (r, j, i, k) , but only the 1-link-loop rule applies will cause us to neglect permutation (j, r, i, k) (Fig. 5.4 (e)). The latter permutation, which has the output r *within* the loop can be evaluated by splitting the loop at i , and then taking the componentwise product of the operators for the two paths to the output r . Since there is no other loop at $\{j, k\}$, we view it as having the input $\mathbf{1}$. (This evaluation of a loop at the output is only necessary if the output is part of the loop.) The result is $(G \odot G^* G^*)a$, which we assume to be small.

Bigger loops (with more inputs a) can be handled by breaking the loop at some node and then forming the appropriate componentwise product operator, giving rise to effective inputs $\{(G \text{ diag } a \ G) \odot G^*\}a$ or $\{G \odot (G^* \text{ diag } a \ G^* \text{ diag } a \ G^*)\}a$ at the combined node. Where there are loops within loops, we recursively evaluate from the innermost loops by replacing each loop with the effective input at the corresponding combined node and working toward the outermost loops. For the purpose of evaluating the high-noise linear, quadratic, and cubic filters, the largest loop will have only two internal inputs, corresponding to, for example, effective inputs $\{(G \text{ diag } a \ G) \odot G^*\}a$ and its reversal $\{G \odot (G^* \text{ diag } a \ G^*)\}a$. The use of approximately self-avoiding processes will allow us to apply the 1-link-loop rule to

neglect these terms because, for example, each column of the “amplified” forward operator $G \text{ diag } a G$ probably has little overlap with the corresponding column of operator G^* . For filters of much higher-order than cubic, the 1-link-loop rule may not be appropriate even for approximately self-avoiding processes because the kernel for a long sequence of alternating forward operators and amplifications (e.g., $G \text{ diag } a G \text{ diag } a \cdots G$) may overlap with G^* ; for example, a may represent an edge map of a sufficiently smooth, closed contour.⁹ Again, we shall apply the 1-link-loop rule for our CIRF Volterra filters up to third order only.

More complicated diagrams than those above can arise; for example, there may be no strict nesting of the loops. However, using the 1-link-loop rule we shall neglect the terms corresponding to such diagrams. These more complex diagrams can represent highly improbable self-intersection patterns—tortuous paths—so neglecting them seems reasonable.

In summary, we use a diagram-based technique for evaluating all the permutations in the terms of the posterior mean expansion (5.14) for CIRF linear, quadratic, and cubic Volterra filters. We exploit the approximate self-avoidingness of our Markov process G to invoke the 1-link-loop rule and the positivity of G for the input-output-coupling rule. As a by-product we neglect a large variety of improbable looping paths. We simplify our work by observing that we get the same diagram after swapping indices in each Kronecker delta, and by implementing permutation reversals by swapping G and G^* appropriately. Unsuppressing η and the appropriate powers of $-\frac{1}{2}$, this diagram-based calculation was used to derive the following.

Result 1 (High-noise MMSE CIRF Volterra Filters). *Suppose that the curve indicator random field U (for approximately self-avoiding Markov processes) is cor-*

⁹For the direction process, if there is no diffusion in direction ($\sigma_\kappa = 0$) and space is unbounded, which corresponds to straight line trajectories, even higher-order kernels will never overlap.

rupted with additive Gaussian white noise of variance $\sigma_N^2 =: \epsilon^{-1}$ to produce measurement vector m . Let $\zeta := d\lambda$, where $d = g^{ii}$ and λ is the average curve length. Then the minimum mean-squared error estimate of U given m has the following approximate asymptotic expansion as $\sigma_N \rightarrow \infty$ ($\epsilon \rightarrow 0$):

Linear Filter:

$$\tilde{u}^{(1)} = \eta\{1 - 2\epsilon\zeta + \epsilon(Gm + G^*m)\} + O(\eta\epsilon^2 + \eta^2\epsilon) \quad (5.16)$$

Quadratic Filter:

$$\begin{aligned} \tilde{u}^{(2)} = & \eta\{1 - 2\epsilon\zeta + 3\epsilon^2\zeta^2 + \epsilon(1 - 2\epsilon\zeta)(Gm + G^*m) \\ & + \epsilon^2(G \text{ diag } m Gm + Gm \odot G^*m + G^* \text{ diag } m G^*m)\} \\ & + O(\eta\epsilon^3 + \eta^2\epsilon) \end{aligned} \quad (5.17)$$

Cubic Filter:

$$\begin{aligned} \tilde{u}^{(3)} = & \eta\{1 - 2\epsilon\zeta + 3\epsilon^2\zeta^2 - 4\epsilon^3\zeta^3 \\ & + \epsilon(1 - 2\epsilon\zeta + 3\epsilon^2\zeta^2)(Gm + G^*m) \\ & + \epsilon^2(1 - 2\epsilon\zeta)(G \text{ diag } m Gm + Gm \odot G^*m + G^* \text{ diag } m G^*m) \\ & + \epsilon^3(G \text{ diag } m G \text{ diag } m Gm + G \text{ diag } m Gm \odot G^*m \\ & + Gm \odot G^* \text{ diag } m G^*m + G^* \text{ diag } m G^* \text{ diag } m G^*m)\} \\ & + O(\eta\epsilon^4 + \eta^2\epsilon). \end{aligned} \quad (5.18)$$

We shall see in Chapter 6 that by implementing the operator G in the Fourier domain, its computational complexity is $O(n \log n)$ for the direction process, where $n = |\mathcal{I}|$. Since multiplying a diagonal matrix with a vector has $O(n)$ complexity, the

net complexity for all of these filters is therefore $O(n \log n)$ for the direction process. This is far better than for the general k -th order Volterra filters, which have $O(n^{k+1})$ complexity.

One could attempt to improve upon these results by pursuing higher-order expansions. However, instead of following that tedious, difficult path, we now turn to a very different approach for deriving filters that is based on the MGF of the CIRF that we derived in Chapter 3. Unlike for general signals, we can follow this path due to the special structure of the CIRF. Not only is the approach simpler, it also allows one to exploit all of the probabilistic structure in the CIRF.

5.7 The Biased CIRF Approximation of the Posterior

In addition to deriving high-noise Volterra filters by using the low-order cumulants of the prior, we can also compute filters by approximating the posterior with a more manageable distribution. This idea has found numerous applications in Markov random fields, Boltzmann machines, and recurrent neural networks, where a highly coupled system is approximated with completely uncoupled system that is trivial to handle [91]. A common theme in those approximations is that the random influences on a site are replaced with their average, giving rise to a coupled system of equations in the unknown means. Inspired by its original application in statistical field theory [90], this procedure is known as the mean-field approximation. More recent presentations of mean-field approximation seek to minimize the Kullback divergence between a given distribution and its approximation by taking a perturbation expansion to a factored distribution [106, 53, 114, 115]. We shall now develop a related approximation for the CIRF, although the approximation will not be factored.

Because by Bayes' rule any posterior is a weighted version of the prior, it is reasonable to base our approximation on the prior. Although the prior itself will not suffice as an approximation, since it is not even a function of the measurements, the prior can be modified or “biased” by a function of the measurements to make a useful approximation of the posterior. To see how to do this, recall from §3.4.2 that we have a tractable expression for the CIRF moment generating functional $\mathbb{E} \exp\langle c, U \rangle$, where the moments are obtained by differentiating the MGF and *evaluating the derivatives at* $c = 0$. However, the MGF of the CIRF converges for *all* c sufficiently small (Prop. 3), and so we could evaluate the derivatives at $c \neq 0$, i.e., perform a Taylor series expansion around $c \neq 0$. Doing so amounts to finding the moments of a CIRF *of bias* c , or *biased* CIRF, with distribution $P_c(du) := Z_c^{-1} \exp\langle c, u \rangle P(du)$, where $c \in \mathbb{R}^{|I|}$ is the bias (vector), $P(du)$ is the original (unbiased) CIRF distribution, and $Z_c := \int \exp\langle c, u \rangle P(du)$. Expectations with respect to P_c are indicated with \mathbb{E}_c . We shall first compute the cumulants of the biased CIRF and then set up a system of equations that allows us to approximate the true posterior with a biased CIRF. All statistics of the posterior can then be approximated by those of the approximating biased CIRF distribution. In particular, the posterior mean of the CIRF can be approximated in this way.

5.7.1 Cumulants of the Biased CIRF

To compute the cumulants of the biased CIRF we need to take derivatives of its cumulant generating functional $\ln \mathbb{E}_c \exp\langle b, U \rangle$ with respect to the vector b , and then evaluate them at 0. Since $\text{CGF}_c(b) := \ln \mathbb{E}_c \exp\langle b, U \rangle = \ln \int \exp\langle b+c, u \rangle P(du) - \ln Z_c$, taking derivatives with respect to b of the CGF of the biased CIRF (denoted $\text{CGF}_c(b)$) and evaluating at 0 is equivalent to taking derivatives with respect to c of the CGF $\ln \mathbb{E} \exp\langle c, U \rangle$ of the (unbiased) CIRF and evaluating at c .

For example, to compute the mean of a biased CIRF, we have to compute

$$\nabla_c \ln \mathbb{E} \exp \langle c, U \rangle = \left(\frac{\partial}{\partial c_1} \ln \mathbb{E} \exp \langle c, U \rangle, \dots, \frac{\partial}{\partial c_{|I|}} \ln \mathbb{E} \exp \langle c, U \rangle \right),$$

where ∇_c denotes the gradient with respect to c . Substituting the CIRF moment generating functional (Prop. 5), we have to take the gradient of $\langle \mu, \bar{\mathcal{N}}(G_c - G)\nu \rangle = \bar{\mathcal{N}}[\langle \mu, G_c \nu \rangle - 1]$, where we recall the notation $G_c := G(c)$ for the Green's operator with bias c . Applying the gradient to the latter inner product, we can write $\nabla_c \langle \mu, G_c \nu \rangle = \langle \mu, [\nabla_c G_c] \nu \rangle$.

Since $G_c := -(Q + \text{diag } c)^{-1}$, to continue the calculation we must understand how to take derivatives of inverses of operators, which we study in the following lemma.

Lemma 4. *Suppose matrices $A = (a_{ij})$ and $B = (b_{ij})$ are functions of some parameter α , i.e., $A = A(\alpha), B = B(\alpha)$. Letting primes ($'$) denote (componentwise) derivatives $d/d\alpha$ with respect to α , the following formulas hold:*

$$(AB)' = A'B + AB', \tag{5.19}$$

$$(A^{-1})' = -A^{-1}A'A^{-1}. \tag{5.20}$$

PROOF. Formula (5.19) follows by applying the single variable product rule to the sum from the inner product of a row of A and column of B : $(AB)'_{ij} = (\sum_k a_{ik} b_{kj})' = \sum_k (a'_{ik} b_{kj} + a_{ik} b'_{kj}) = (A'B + AB')_{ij}$.

To obtain formula (5.20), we first apply formula (5.19) to write $0 = I' = (AA^{-1})' = A'A^{-1} + AA^{-1}'$. Then we solve for A^{-1}' . \square

Returning to the calculation of the mean of the biased CIRF, we use (5.20) to write $\frac{\partial}{\partial c_i} G_c = \frac{\partial}{\partial c_i} [-(Q + \text{diag } c)^{-1}] = (Q + \text{diag } c)^{-1} \left[\frac{\partial}{\partial c_i} (Q + \text{diag } c) \right] (Q + \text{diag } c)^{-1} = G_c D_i G_c$, where D_i is a matrix with a one at the (i, i) -entry and zeroes elsewhere. The

matrix D_i can be expressed as the outer product $\delta_i \delta_i^*$ of discrete impulse vectors δ_i at site i , where $(\delta_i)_j = \delta_{ij}$. Therefore $\frac{\partial}{\partial c_i} \langle \mu, G_c \nu \rangle = (\delta_i^* G_c^* \mu) (\delta_i^* G_c \nu) = (G_c^* \mu)_i (G_c \nu)_i$, proving the following:

Proposition 7. *The mean of the CIRF of bias c at site i is*

$$\kappa_c^i = \mathbb{E}_c U_i = \bar{\mathcal{N}} \langle \mu, G_c D_i G_c \nu \rangle. \quad (5.21)$$

Equivalently, the mean (vector) of this biased CIRF is

$$\mathbb{E}_c U = \bar{\mathcal{N}}(G_c^* \mu) \odot (G_c \nu). \quad (5.22)$$

The higher-order cumulants can also be tractably characterized as follows:

Proposition 8. *The k -th cumulant at sites i_1, \dots, i_k of the CIRF U of bias c is*

$$\begin{aligned} \kappa_c^{i_1, \dots, i_k} = \text{cum}_c \{U_{i_1}, \dots, U_{i_k}\} &= \bar{\mathcal{N}} \langle \mu, \sum_{a \in \text{perm}_k} G_c D_{i_{a_1}} G_c \cdots G_c D_{i_{a_k}} G_c \nu \rangle \\ &= \bar{\mathcal{N}} \sum_{i, j} \mu_i \nu_j \sum_{a \in \text{perm}_k} g_c^{i i_{a_1}} g_c^{i_{a_1} i_{a_2}} \cdots g_c^{i_{a_{k-1}} i_{a_k}} g_c^{i_{a_k} j}, \end{aligned} \quad (5.23)$$

where $(g_c^{ij}) := G_c$ is the Markov process Green's operator with bias c .

PROOF. Since by definition $\text{cum}_c \{U_{i_1}, \dots, U_{i_k}\} = \frac{\partial^k}{\partial c_{i_1} \cdots \partial c_{i_k}} (\bar{\mathcal{N}} \langle \mu, G_c \nu \rangle - \bar{\mathcal{N}})$, we need to show

$$\frac{\partial^k G_c}{\partial c_{i_1} \cdots \partial c_{i_k}} = \sum_{a \in \text{perm}_k} G_c D_{i_{a_1}} G_c \cdots G_c D_{i_{a_k}} G_c. \quad (5.24)$$

(Here and in the following, all derivatives will be evaluated at c .) We shall prove (5.24) by induction. By the proof of Prop. 7, formula (5.24) holds for $k = 1$. Suppose

(5.24) holds for $k = n$. We now check the case $k = n + 1$:

$$\begin{aligned}
\frac{\partial^{n+1} G_c}{\partial c_{i_1} \cdots \partial c_{i_{n+1}}} &= \frac{\partial}{\partial c_{i_{n+1}}} \sum_{a \in \text{perm}_n} G_c D_{i_{a_1}} G_c \cdots G_c D_{i_{a_n}} G_c \\
&= \sum_{j=1}^{n+1} \sum_{a \in \text{perm}_n} G_c D_{i_{a_1}} G_c \cdots G_c D_{i_{a_{j-1}}} (G_c D_{i_{n+1}} G_c) D_{i_{a_j}} G_c \cdots G_c D_{i_{a_n}} G_c \\
&= \sum_{a \in \text{perm}_{n+1}} G_c D_{i_{a_1}} G_c \cdots G_c D_{i_{a_{n+1}}} G_c,
\end{aligned}$$

using $\frac{\partial G_c}{\partial c_i} = G_c D_i G_c$ from the proof of (5.21) again. The result follows. \square

5.7.2 Solving for the Biased CIRF

Given this manageable characterization of the cumulants of the biased CIRF, our task now is to solve for the bias c to approximate the posterior. Because the logarithm $\ln Z_{(\epsilon)}$ of the normalizing constant $Z_{(\epsilon)}$ of the posterior is (up to an additive constant) just the cumulant generating functional of the posterior (see §5.7.1), we shall develop an approximation to $\ln Z_{(\epsilon)}$. The general strategy is similar to that for the Volterra filter, where we assume there is high-noise (ϵ near 0), and so a perturbation expansion in ϵ is warranted. Except in the sense of the remainder term of the expansion, we do not provide approximation bounds¹⁰. Letting $\epsilon c' = c$, we write

$$\begin{aligned}
\ln Z_{(\epsilon)} &= \ln \int e^{\epsilon \langle a + Au, u \rangle} P(du) \\
&= \ln \int e^{\epsilon \langle a - c' + Au, u \rangle} e^{\epsilon \langle c', u \rangle} P(du) \\
&= \ln Z_{\epsilon c'} \mathbb{E}_{\epsilon c'} e^{\epsilon \langle a - c' + AU, U \rangle} \\
&= \ln Z_c + \ln \mathbb{E}_c e^{\epsilon \langle a - c' + AU, U \rangle}, \tag{5.25}
\end{aligned}$$

¹⁰Approximation bounds for sigmoidal belief networks are provided in [106].

where the second term on the last line is simply the cumulant generating functional of $\langle a - c' + AU, U \rangle$ (with respect to measure $P_c(du)$) evaluated at ϵ , similar to (5.11) except the measure on U is now biased by c . Now, as for the Volterra filters, we shall develop the CGF of $\langle a - c' + AU, U \rangle$ into a Taylor series in ϵ . Unlike the Volterra case, however, we have the extra freedom to choose the bias c to make our approximation useful, and therefore we shall stop the Taylor series at the first order, i.e.,

$$\ln \mathbb{E}_c e^{\epsilon \langle a - c' + AU, U \rangle} = \ln(1 + \mathbb{E}_c \langle a - c' + AU, U \rangle + O(\epsilon^2) = \mathbb{E}_c \langle a - c' + AU, U \rangle + O(\epsilon^2).$$

Because the latter expectation is just $-\langle c, \mathbb{E}_c U \rangle + \epsilon \langle a, \mathbb{E}_c U \rangle + \epsilon \sum_j \mathbb{E}_c[(AU)_j U_j]$, we can rewrite (5.25) as

$$\ln Z_{(\epsilon)} = \ln Z_c - \langle c, \mathbb{E}_c U \rangle + \epsilon \left\{ \langle a, \mathbb{E}_c U \rangle + \sum_j [\text{cov}_c((AU)_j, U_j) + \mathbb{E}_c(AU)_j \mathbb{E}_c U_j] \right\} + O(\epsilon^2). \quad (5.26)$$

Since $\ln Z_{(\epsilon)}$, which characterizes the true posterior, does not depend on the bias c of our approximate posterior P_c , differentiating $\ln Z_{(\epsilon)}$ with respect to any function of c will be zero. In particular, we can set up a system of equations in c by taking the derivative of (5.26) with respect to $\mathbb{E}_c U_i = \kappa_c^i$, a function of c . We consider two special cases now.

No Blur Case

Consider the situation where there is no blur in the observations and the noise is white. Thus we set $A = -\frac{1}{2}I = -\frac{1}{2}(\delta_{ij})$ and (5.26) becomes

$$\ln Z_{(\epsilon)} = \ln Z_c - \sum_j c_j \kappa_c^j + \epsilon \left\{ \sum_j a_j \kappa_c^j - \frac{1}{2} \sum_j [\kappa_c^{j,j} + \kappa_c^j \kappa_c^j] \right\} + O(\epsilon^2). \quad (5.27)$$

Differentiating with respect to κ_c^i and applying the chain rule twice, the first two terms on the right hand side become

$$\frac{\partial}{\partial \kappa_c^i} \left[\ln Z_c - \sum_j c_j \kappa_c^j \right] = \sum_j \frac{\partial \ln Z_c}{\partial c_j} \frac{\partial c_j}{\partial \kappa_c^i} - c_i - \sum_j \frac{\partial c_j}{\partial \kappa_c^i} \kappa_c^j = -c_i, \quad (5.28)$$

because $\frac{\partial \ln Z_c}{\partial c_j} = \kappa_c^j$. Using formula (5.23) twice, note that $\kappa_c^{jj} = 2g_c^{jj} \bar{\mathcal{N}} \sum_{i', j'} \mu_{i'} \nu_{j'} g_c^{i'j} g_c^{jj'} = 2g_c^{jj} \kappa_c^j$, because the two permutations are identical. So by differentiating (5.26) with respect to κ_c^i , we get: $0 = -c_i + \epsilon(a_i - g_c^{ii} + \mathbb{E}_c U_i) + O(\epsilon^2)$, proving the following result:

Result 2. *Suppose the posterior distribution of the observed CIRF is $P_{(\epsilon)}(du) \propto \exp(\epsilon(a - \frac{1}{2}u, u))P(du)$. Then $P_{(\epsilon)}$ can be approximated by a CIRF of bias c , with*

$$c = \epsilon[a - \mathbb{E}_c U - d(c)] + O(\epsilon^2), \quad (5.29)$$

where $d(c)$ is the diagonal of the Green's operator G_c with bias c , or $d(c)_i := g_c^{ii}$.

As discussed in §5.2.1, we can apply this result to the case where the CIRF is corrupted with additive Gaussian white noise by setting $a = m$ and $\epsilon = \sigma_N^{-2}$. This can also be applied when empirical edge histograms are known (§5.2.3) or in other situations by setting a to an appropriate function of m .

Observe that (5.29) is a coupled nonlinear system in the vector c of $|\mathcal{I}|$ unknowns; c is expressed as a function not only of the exogenous a , but also of the (generally nonlinear) functions of itself: $\mathbb{E}_c U$ (see Proposition 7) and $d(c)$. This equation describes c as proportional to a balance between positive influence due to input a and negative influence due to the sum of the mean of the biased CIRF ($\mathbb{E}_c U$) and a measure of closure $d(c)$. Recall that $d(c)_i = (G_c)^{ii}$ is the expected amount of time that the biased Markov process spends in site i given that it started at i . We note that

even if the unbiased process is self-avoiding, the bias may be such as to encourage the process to return to i . In the context of Mumford's direction process, if c is sufficiently positive along an appropriate dilation of a smooth closed curve passing through i in discrete $\mathbb{R}^2 \times \mathbb{S}$, then $d(c)_i$ may be large. This closure measure may not be significant if the bias describes only open curves. Since the bias c is ultimately a function of input a , we would expect $d(c)$ to be important only when a is positive along smooth, closed curves.¹¹

Large Blur Case

Another system of equations can be derived in an opposite extreme, where the symmetric negative definite matrix A is proportional to an operator that performs a great deal of blur. In that case, the blurred field AU is essentially a constant vector, and therefore $\text{cov}_c((AU)_j, U_j) \approx 0$. Using $\nabla_{\mathbb{E}_c U} \langle \mathbb{E}_c AU, \mathbb{E}_c U \rangle = 2A\mathbb{E}_c U$ and the above calculations for the no-blur case, we get the following system of equations.

Result 3. *Suppose the posterior distribution of the observed CIRF is $P_{(\epsilon)}(du) \propto \exp(\epsilon \langle a + Au, u \rangle) P(du)$, where the symmetric negative definite operator A is approximately proportional to a spatial averaging operator of maximum size, such as a matrix of constants. Then $P_{(\epsilon)}$ can be approximated by a CIRF of bias c , with*

$$c \approx \epsilon[a + 2A\mathbb{E}_c U] + O(\epsilon^2). \quad (5.30)$$

This large blur result provides a nonlinear system of equations in the bias c . It differs from (5.29) by the presence of the blur operator A acting on the mean and

¹¹The exact notion of smoothness and closedness will depend on the Markov process curve model. For the direction process, closed curves with curvature not many times that of σ_κ might induce a $d(c)$ having large values at some sites.

the absence of the closure measure $d(c)$. By including blur, we can better describe a number of situations, as outlined in §5.2

Also note that for both the non-blur and large blur cases, the biased CIRF approximations do not depend on a spatial homogeneity assumption, and so nonuniform μ and ν can be used.

Combined with the formula for the mean of the biased CIRF and dropping terms of order ϵ^2 , our nonlinear system corresponding to (5.30) can be expressed as the following coupled nonlinear system describing a self-consistent field c mediating the posterior CIRF interactions:

Observation (Biased CIRF Approximation of the Posterior Mean of the CIRF). *Suppose a CIRF U is corrupted via a quadratic log-likelihood (with large-scale blur) to produce measurements m . The posterior mean of U given m has the following approximate description:*

$$\mathbb{E}_m U \approx \bar{\mathcal{N}} f \odot b, \quad (5.31)$$

where f and b solve the following coupled system:

$$Qf + \epsilon \text{diag}[a + 2\bar{\mathcal{N}}A(f \odot b)]f + \nu = 0 \quad (5.32)$$

$$Q^*b + \epsilon \text{diag}[a + 2\bar{\mathcal{N}}A(f \odot b)]b + \mu = 0. \quad (5.33)$$

For reference, we include the above filtering equations for the continuum-limit of the above biased CIRF approximation equations for the direction process.

Observation (Reaction-Diffusion-Convection Equation for the Direction Process). *Suppose a CIRF U based on the direction process is corrupted via a quadratic log-likelihood (with large-scale blur) to produce measurements m . Then the posterior*

mean is approximately $\bar{\mathcal{N}}f \odot b$, where f and b solve the following coupled nonlinear integro-elliptic system:

$$\begin{aligned} \left(\frac{\sigma^2}{2} \frac{\partial^2}{\partial \theta^2} - \cos \theta \frac{\partial}{\partial x} - \sin \theta \frac{\partial}{\partial y} - \alpha \right) f + \epsilon \operatorname{diag}[a + 2\bar{\mathcal{N}}A(f \odot b)]f + \nu &= 0 \\ \left(\frac{\sigma^2}{2} \frac{\partial^2}{\partial \theta^2} + \cos \theta \frac{\partial}{\partial x} + \sin \theta \frac{\partial}{\partial y} - \alpha \right) b + \epsilon \operatorname{diag}[a + 2\bar{\mathcal{N}}A(f \odot b)]b + \mu &= 0. \end{aligned}$$

Also note that for A , the nonlinear coupled PDE is in fact a reaction-diffusion-convection equation. Since $c \approx \epsilon[a + 2\bar{\mathcal{N}}A(f \odot b)]$, we see that pattern formation arises from $c^+ = \max(c, 0)$ (the positive part of bias c), pattern repression comes from $-c^- = \min(c, 0)$ (the negative part of c), diffusion comes from the second derivatives, and convection is due to the first derivatives [136].

5.8 Comparison with Completion Fields

Perceptual organization has tantalized a number of researchers in computer vision with the possibility of discovering a rigorous foundation for classical demonstrations such as illusory contours [52]. This gave rise to work on contour saliency both as a model of illusory contour formation and for contour organization. We shall focus on drawing relationships to stochastic completion fields [126, 109] and eigensources [117] because of their focus on Markov processes such as the direction process. Stochastic completion fields in particular were a strong inspiration for this thesis.

To make the connection, we begin with the biased CIRF approximation, but set the operator A to zero. The first consequence of dropping A is that the likelihood is no longer necessarily normalizable. For the second consequence, recall the interpretation of the square of the CIRF U as a measure of contour intersections (see §3.1.1). Ignoring the operator A (which gives rise to a quadratic form in U) therefore amounts to

ignoring those intersections (due to the diagonal of A) or even near intersections (due to the off-diagonals of A). With this simplification this biased CIRF approximation becomes the decoupled forward and backward linear systems $Qf + Cf + \nu = 0$ and $Q^*b + Cb + \mu = 0$, respectively, where $C := \epsilon \text{diag } a$. The approximate posterior mean of the CIRF U is proportional to the componentwise product $f \odot b$ of the forward and backward solutions. We focus on the forward equation, as the backward equation is just its adjoint.

Now, we observe even though we have dropped the interaction operator A , the resulting filter is still a nonlinear input-to-output mapping, and in two distinct ways. First, we are taking the product of the forward and backward solutions. Second, the input arises as a modification to the diagonal of linear operator Q . In particular, we have $f = (I - GC)^{-1}G\nu = \sum_{k=0}^{\infty}(GC)^k G\nu$, which says that f is an infinite polynomial in c , i.e., f is nonlinear in c . If we ignore the constant term ($k = 0$) and all nonlinear terms ($k > 1$), what remains is the linear term $GCG\nu$, which is proportional to Gc if ν is a constant vector and G is spatially homogeneous (as it is for the direction process). We recognize Gc as the *source field* (with sources $c \propto CG\nu$), and, by the same argument, G^*c is the *sink field* (with sinks $c \propto CG^*\mu$, where μ is also a constant vector); their product $Gc \odot G^*c$ is then Williams' and Jacobs' stochastic completion field [126]. To have different sinks and sources one can adjust μ and ν .

To relate to eigensources, let $C := \text{diag } a$ and thus write the forward equation as: $f - \epsilon GCf - G\nu = 0$. Setting $\nu = 0$ and assuming $a_i > 0$, for all i , we can define $W^2 := C$ to get $GW^2f = \sigma_N^2 f$, where we recall $\epsilon = \sigma_N^{-2}$ (see §5.2.2). Defining $h = Wf$, we write $GW h = \sigma_N^2 W^{-1}h$, or

$$WGW h = \sigma_N^2 h.$$

This is an eigenvalue equation of the same form as in [117], where σ_N^2 is now interpreted as an eigenvalue of WGW . We see, therefore, that the CIRF provides a new perspective for understanding other techniques for the perceptual organization of contours.

5.9 What Can One Do with a Filtered CIRF?

It may seem odd that we have sought to (merely) filter a field of local contour measurements, producing yet another field. If one objective of vision is to explicitly represent structures with increasing abstraction, this task may seem to postpone the most difficult work. To return to a question brought up at the start of this chapter, why not extract the contours directly from the image, instead of just enhancing a contour image?

There are a number of reasons why one might focus on processing a field. First, this approach captures a notion of global information beyond that of a single curve, by explicitly expressing the idea of an *image of* contours, where both the interactions among the contours and the blur due to imaging is not ignored. In contrast, previous approaches for computing curves using Markov process contour models [80, 108, 35, 15, 26, 126, 133], implicitly assume that there is only a single contour. By emphasizing the field we can combine the richer contour models of the single-contour schemes with the contour interactions implicit in the line processes used in MRFs [36] or dictionary-based relaxation labeling [40]. Another reason for the intermediate representation of a filtered CIRF is that it captures the ambiguity inherent in finding contours in noisy images: a more diffuse filter response characterizes less certain (more spread out) contours. If direct contour computations were carried out, one would need a set of contours to represent this ambiguity. While in some appli-

cations only the best contour is needed, in others we require an assessment of the uncertainty of the situation and some representation of the plurality of comparable contour explanations.¹² Direct enumeration with particular contours may be more cumbersome than a field for representing the many contour possibilities. Finally, contour enhancement can be an end in itself in some graphical or image processing applications.

More pragmatically, simple post-processing of the filtered CIRF may be sufficient for some applications. For example, the thresholding described in §5.1.4 may be sufficient, perhaps followed by connected component analysis (in 3-d for the directed Mumford process, for example) to provide a thick individual contours, providing a description qualitatively similar to Iverson’s thick trace [49]. Thinning and linking procedures could then follow if desired.

Since information is lost in thresholding, however, for more sophisticated contour processing, might require the (un-thresholded) CIRF posterior mean. To compute contours one could greedily follow gradients in the posterior mean from strong, supra-threshold seed points. The CIRF posterior mean could also be used to guide a heuristic search procedure for finding individual contours [73]. Since the (filtered) CIRF is the expected amount of time a contour spends at a site (given the measurements), the posterior mean is a candidate heuristic function that would encourage sites with large times to be explored first and more often than others. Other graph-based methods [12] could be used to post-process the CIRF in a similar fashion. By combining information about the observation process and the structure of contours, the CIRF performs significant global computations, and so reduces the burden upon later processing.

¹²See [116] for explicit calculations of such contour diffusivity.

Chapter 6

Numerical Methods

Having worked out formulas for MMSE CIRF filters, we now discuss how these can be implemented. While the number of linear and nonlinear filters we have suggested may seem daunting, on closer inspection of the formulas in Chapter 5 we see some common components. For example, the Green's operator G for the generator L of the Markov process appears in the linear and Volterra filters. In addition, the Green's operator G_c of the biased CIRF also appears prominently in filters based on the biased CIRF approximation of the CIRF posterior. We therefore first study the problem of computing the action $f = Gh$ of G upon h for both the direction and curvature processes, where in both cases we have direct (noniterative) methods. Knowing how to apply G to a vector enables us to compute the action of the biased Green's operator G_c on a vector using an iterative technique that uses G as a preconditioner. The technique for computing G_c does not itself depend upon the Markov process because those details are summarized by G . Finally, we present a continuation method for solving the nonlinear system for the biased CIRF. The continuation method requires that we be able to apply G_c , and so it is also not tied to a particular Markov process.

6.1 Applying the Green's Operator by Solving the Forward Equation

To compute the action f of the Green's operator G upon a vector h , we can—equivalently—solve for f in the *forward equation*

$$(\alpha I - L)f = h, \tag{6.1}$$

since $Q = L - \alpha I$ and $G = -Q^{-1}$. Now, all of the theory in Chapters 3 and 5 was based on continuous-time, *discrete*-space Markov processes, to avoid many measure-theoretic distractions. However, as we saw in Chapter 2, the Markov processes we are most interested in live in *continuous* spaces such as $\mathbb{R}^2 \times \mathbb{S}$. In general, by designing a Markov process via an appropriate stochastic differential equation, we can immediately obtain the process's generator \mathcal{L} from the corresponding Fokker-Planck equation. When the Markov process is a diffusion process, \mathcal{L} is a partial differential operator acting on functions of continuous space, not vectors. To make the transition from the continuous to discrete domains, we let the discrete-space generator L be some (reasonable) discretization of the continuous-space generator \mathcal{L} . Strictly speaking, therefore, the CIRF filters are only fully defined when we specify the numerical method used to obtain a particular (discrete) generator L . We discretize only after having studied the filtering problem to focus on the underlying mathematics of filtering curve images, not the (often crucial) details of implementation. By this thinking, we may imagine that our filters require the solution of differential or integral equations, even though we only solve their discretizations. In addition, although we employ particular numerical techniques here, we note that they are not unique, as in many numerical problems.

6.1.1 Direction Process

Here we solve the (discrete) forward equation for Mumford's direction process by first specifying the related direction process forward PDE and then discretizing using Fourier series. The result will be a set of linear systems in the frequency domain that can be solved rapidly. After taking an inverse discrete Fourier transform, we shall find that we can accurately compute the action $f = Gh$ of G on a vector $h \in \mathbb{R}^n$ in $O(n \log n)$ -time. Because this technique does not require iteration, we call it the *direct method*.

Recalling the continuous-space generator \mathcal{L} of the direction process from (2.4), the corresponding forward PDE is the continuous space equivalent $(\alpha I - \mathcal{L})f = g$ of (6.1), or the *direction process forward PDE*

$$\alpha f - \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial \theta^2} + \cos \theta \frac{\partial f}{\partial x} + \sin \theta \frac{\partial f}{\partial y} = h, \quad (6.2)$$

where $f = f(x, y, \theta)$, $h = h(x, y, \theta)$, and $\sigma = \sigma_\kappa > 0$. Observe that this is a degenerate elliptic PDE, where the degeneracy arises because the second derivatives do not appear in x and y . As Mumford [82] and later Williams and Jacobs [125] have noted for the corresponding Fokker-Planck diffusion equation, the second derivative in θ gives rise to diffusion in the θ -dimension, and the cosine/sine pair of terms give rise to convection. Suppose we are to solve (6.2) on $\mathcal{D} \times [0, 2\pi]$, for planar rectangular region $\mathcal{D} = [0, x_{\max}] \times [0, y_{\max}] \subset \mathbb{R}^2$. Let $\partial\mathcal{D}$ denote the boundary of \mathcal{D} . For boundary conditions, we always assume $f(x, y, \theta)$ is periodic in θ with period 2π . With respect to (x, y) , one may study several possible boundary conditions:

- *Dirichlet*: $f(\partial\mathcal{D}, \theta) = 0, \quad \forall \theta;$
- *Neumann*: $\frac{\partial f}{\partial \mathbf{n}}(\partial\mathcal{D}, \theta) = 0, \quad \forall \theta, \quad \text{where } \mathbf{n} \text{ is the outward normal of } \mathcal{D};$

- *Periodic:* f is periodic in x with period x_{\max} and in y with period y_{\max} .

The Dirichlet and Neumann boundary conditions may be more natural for contour enhancement, as periodic boundary conditions can lead to “wrap around” artifacts.

At least two major approaches can be taken in numerically solving PDE (6.2): (1) finite differencing, or (2) projection onto another basis. In the first approach, the partial derivatives are replaced with (discrete) difference approximations, similar to the definition of a derivative except without taking the limit. Finite differencing not only is simple to implement, but it also allows boundary conditions to be easily specified and modified. Williams and Jacobs used finite differencing to solve the corresponding time-dependent Fokker-Planck PDE (with decay), where the approximation was only first-order accurate because upwind differencing was needed for stability [125]. They observed a troubling anisotropy with this first-order method: the solution is more focused in the horizontal and vertical directions than in the diagonals (see Fig. 6.1). In other words, rotating h does not simply rotate the solution f , as PDE (6.2) would predict. Extra diffusion was later added to restore isotropy [127], with the additional benefit of enabling a fast multi-grid solution. Unfortunately, this technique ensured that all directions had the same unfocused response as the diagonals did without extra diffusion. It has been difficult to achieve accuracy in solving the direction process forward PDE while maintaining isotropy using finite differencing.

A second approach to solving the direction process forward PDE (6.2) involves selecting a suitable (and typically orthogonal) basis $\{\phi_k\}$, taking the inner product $\langle \phi, \psi \rangle_{\mathcal{R}} := \int \phi(r) \psi^*(r) dr$ (where the asterisk denotes complex conjugation, $r = (x, y, \theta) \in \mathcal{R} = \mathbb{R}^2 \times \mathbb{S}$, and the subscript \mathcal{R} is dropped if understood from the context) of the equation with the basis elements, and then solving the transformed system. A judicious selection of basis can lead to reduced computation times and/or solution quality. One important virtue of the complex exponential

(or complex Fourier) basis is that derivatives can be represented exactly by scalar multiplication, suggesting that the anisotropy introduced by finite differencing might be alleviated. In addition, the ready availability of reliable implementations of the fast Fourier transform (FFT), which compute the projection to the complex exponential basis, make coding straightforward. Finally, we shall see that PDE (6.2) reduces to a set of cyclic tridiagonal systems, which can be rapidly solved. Zweck and Williams solved the related time-dependent diffusion equation by projecting onto localized wavelet-like bases and obtained superb isotropic performance; unfortunately, the computational cost of their scheme is greater than the cost of finite differencing [140]. The need to apply the Green's operator possibly many times to implement our MMSE filters motivated us to devise the following rapid scheme.

Before developing our method, we observe that trigonometric bases correspond to the three different boundary conditions. Specifically, the sine, cosine, and complex exponential bases correspond to the Dirichlet, Neumann, and periodic boundary conditions respectively. Despite the periodicity of the complex exponential basis, we apply it here both because of the above advantages and to simplify the presentation. To reduce the negative artifacts of periodicity in (x, y) , we can often solve the equation on a larger domain than necessary (after suitably padding h), and then extract the region of interest from the solution f .

We now detail our direct method. Let \mathfrak{i} denote the imaginary unit $\sqrt{-1}$. The *complex exponential* having periods x_{\max} , y_{\max} , and 2π and at frequencies j , k , and l in the x -, y -, and θ -dimensions, respectively, is $\phi_{jkl}(x, y, \theta) := \exp \left[2\pi\mathfrak{i} \left(\frac{jx}{x_{\max}} + \frac{ky}{y_{\max}} + \frac{l\theta}{2\pi} \right) \right]$. Let the Fourier coefficient \hat{f}_{jkl} be defined as the inner product $\langle f, \phi_{jkl} \rangle$ of a function f with a complex exponential ϕ_{jkl} ; \hat{h}_{jkl} is defined similarly. Now we can evaluate, term-by-term, the inner product of the complex exponential with the direction process forward PDE.

Using integration by parts and periodicity in x , we know $\langle \frac{\partial f}{\partial x}, g \rangle = -\langle f, \frac{\partial g}{\partial x} \rangle$; this result (“taking the adjoint of the x -derivative”) holds analogously in y and in θ . By both taking the adjoint of the θ -derivative twice and applying the chain rule twice, we get $\langle \frac{\partial^2 f}{\partial \theta^2}, \phi_{jkl} \rangle = (-1)^2 \langle f, \frac{\partial^2 \phi_{jkl}}{\partial \theta^2} \rangle = (-\mathbb{I}l)^2 \langle f, \phi_{jkl} \rangle = -l^2 \hat{f}_{jkl}$. Using the formula $\cos \theta = (e^{\mathbb{I}\theta} + e^{-\mathbb{I}\theta})/2$ and taking the adjoint of the x -derivative, we can write $\langle \cos \theta \frac{\partial f}{\partial x}, \phi_{jkl} \rangle$ as

$$\begin{aligned} & -\frac{2\pi \mathbb{I}j}{x_{\max}} \frac{1}{2} \int \int \int f(x, y, \theta) \{e^{\mathbb{I}\theta} + e^{-\mathbb{I}\theta}\} e^{-2\pi \mathbb{I}(\frac{jx}{x_{\max}} + \frac{ky}{y_{\max}} + \frac{l\theta}{2\pi})} dx dy d\theta \\ &= \pi \frac{\mathbb{I}j}{x_{\max}} \int \int \int f(x, y, \theta) \left\{ e^{-2\pi \mathbb{I}(\frac{jx}{x_{\max}} + \frac{ky}{y_{\max}} + \frac{(l-1)\theta}{2\pi})} + e^{-2\pi \mathbb{I}(\frac{jx}{x_{\max}} + \frac{ky}{y_{\max}} + \frac{(l+1)\theta}{2\pi})} \right\} dx dy d\theta \\ &= \pi \frac{\mathbb{I}j}{x_{\max}} (\hat{f}_{j,k,l-1} + \hat{f}_{j,k,l+1}). \end{aligned}$$

Similarly, we get $\langle \sin \theta \frac{\partial f}{\partial y}, \phi_{jkl} \rangle = \pi \frac{k}{y_{\max}} (\hat{f}_{j,k,l-1} - \hat{f}_{j,k,l+1})$, using the formula $\sin \theta = (e^{\mathbb{I}\theta} - e^{-\mathbb{I}\theta})/(2\mathbb{I})$. This proves the following.

Proposition 9. *Suppose $h(x, y, \theta)$ is periodic in x , y , and θ with periods x_{\max} , y_{\max} , and 2π , respectively. Then the solution $f = f(x, y, \theta)$ of (6.2) is also periodic and, for each $j, k \in \mathbb{Z}$, the Fourier coefficients $\{\hat{f}_{jkl} : l \in \mathbb{Z}\}$ and $\{\hat{h}_{jkl} : l \in \mathbb{Z}\}$ satisfy the following tridiagonal linear system indexed by $l \in \mathbb{Z}$:*

$$\pi \left(\frac{\mathbb{I}j}{x_{\max}} + \frac{k}{y_{\max}} \right) \hat{f}_{j,k,l-1} + \left(\frac{\sigma^2}{2} l^2 + \alpha \right) \hat{f}_{j,k,l} + \pi \left(\frac{\mathbb{I}j}{x_{\max}} - \frac{k}{y_{\max}} \right) \hat{f}_{j,k,l+1} = \hat{h}_{j,k,l}. \quad (6.3)$$

To go from the above *infinite* linear system to a manageable finite one, we assume that h is bandlimited. Then, without loss, we can represent h by its uniform samples $h_{j'k'l'} := h(j'\Delta x, k'\Delta y, l'\Delta\theta)$, with indices $j' \in \{0, \dots, J-1\}$, $k' \in \{0, \dots, K-1\}$, and $l' \in \{0, \dots, L-1\}$, where positive integers J, K , and L are the total¹ numbers of frequencies in x , y , and θ , respectively, such that the periods satisfy $x_{\max} =$

¹Both nonnegative and negative frequencies are included in this total.

$J\Delta x, y_{\max} = K\Delta y$, and $2\pi = L\Delta\theta$. Prop. 9 then implies f is bandlimited in x and y . Due to the bandlimitedness of h in θ and the fact that the direction process forward PDE is a diffusion in θ (and thus low-pass), we elect to solve for \hat{f} in (6.3) only for the same L total frequencies in θ that are in the bandlimit of h . To deal with the “boundary conditions” in frequency l , we assume that \hat{f}_{jkl} is periodic in l because sampled signals have periodic spectra.² As a result, we obtain an $L \times L$ *cyclic tridiagonal* linear system at each (j, k) frequency pair. Assuming L is even, this sytem is $A\hat{f}_{j,k,\cdot} = \hat{h}_{j,k,\cdot}$, where the matrix A is:

$$\begin{bmatrix} \gamma_{-L/2+1} & \beta^- & 0 & \cdots & \cdots & 0 & \beta^+ \\ \beta^+ & \ddots & \ddots & \ddots & & & 0 \\ 0 & \ddots & \gamma_{-1} & \beta^- & 0 & & \vdots \\ \vdots & \ddots & \beta^+ & \gamma_0 & \beta^- & \ddots & \vdots \\ \vdots & & 0 & \beta^+ & \gamma_1 & \ddots & 0 \\ 0 & & & \ddots & \ddots & \ddots & \beta^- \\ \beta^- & 0 & \cdots & \cdots & 0 & \beta^+ & \gamma_{L/2} \end{bmatrix}, \quad (6.4)$$

where $\gamma_l = \frac{\sigma^2}{2}l^2 + \alpha$, $\beta^+ = \pi \left(\frac{j}{J\Delta x} + \frac{k}{K\Delta y} \right)$, and $\beta^- = \pi \left(\frac{j}{J\Delta x} - \frac{k}{K\Delta y} \right)$. This system can be solved for $\hat{f}_{j,k,\cdot}$ in $O(L)$ operations [96]. Using the FFT to compute the (finite number of) Fourier coefficients of h and the inverse FFT to compute f from \hat{f} , the overall time complexity of computing f from h is $O(n \log n)$, where $n = J \cdot K \cdot L$.

When the input—before discretization—is not bandlimited, aliasing and oscillation can occur both in the (discrete) input and solution. For example, if the discretization of h is a discrete impulse (a “1” at one position, with “0”s elsewhere), the corresponding continuous-space h is *not* a Dirac δ -distribution, since

²Periodic boundary conditions in l produced more pleasing results than zero boundary conditions in an informal test where the θ -dimension was badly undersampled. If the θ -dimension is sufficiently oversampled, the difference is negligible.

the δ -distribution is not band-limited. Instead, the continuous-space h corresponding to a discrete impulse is a sinc function³ centered on the impulse and with zero crossings at all other sample locations. *Between* the sample locations the sinc oscillates above and below zero. If the discrete impulse is shifted by a non-integral number of samples, which can only fairly mean that the underlying continuous signal—the sinc function—is so shifted and then resampled, the discretization will no longer be a discrete impulse, but will have (discrete) oscillations. This phenomenon also appears in the solution of the direction process forward PDE (Fig. 6.2).

The corresponding backward equation $(\alpha I - L^*)f = h$ requires taking the adjoint of the direction process generator, which in this case amounts to flipping the signs on the first derivatives. Obeying these sign changes, the backward equations can be solved similarly.

Instead of numerically solving the direction process forward PDE, one can also compute the action $f = Gh$ by particle simulation [126]. This technique amounts to the estimation of the integral kernel $g(x_0, y_0, \theta_0; x_1, y_1, \theta_1)$ for the continuous-space operator \mathcal{G} (used to idealize $G = (g_{ij})$) directly by simulating a large number of particles behaving as the direction process, and then taking advantage of the interpretation that g_{ij} is the average amount of time the particles spent in location j , given that they started at i , where i and j are discrete-space approximations of (x_0, y_0, θ_0) and (x_1, y_1, θ_1) , respectively. Error will be introduced both in the time discretization required to track the individual particles on a digital computer and in the spatial discretization of the Markov process. The action $f = Gh$ can be computed by explicit integration, leading to an $O(n^2)$ -time computation. Not only are such integrations expensive, the particle simulations can be as well (to ensure that

³The (standard) sinc function has a period of infinity. For finite periods, qualitatively similar functions arise.

the averages have low variance⁴). To avoid the error due to the finite number of particles, Thornber and Williams also set up related diffusion processes in direction and scale for which convenient approximate formulas for their Green's function kernels were obtained analytically using the method of steepest descent [116, 118].

Results

To demonstrate the performance of the direct method we compared it with the first-order finite differencing method. To solve the direction process forward PDE with a first-order method, one can solve the corresponding (time-dependent) Fokker-Planck equation and then numerically integrate over time [125]. Instead, we avoid the explicit time parameter by first analytically integrating over time and then applying the (unpreconditioned) conjugate method on the normal equations (see §6.2), using the same upwind differencing for the spatial variables as in [125].

In the first experiment, we illustrate the isotropy that results from the accuracy of the direct method, using a slightly blurred⁵ impulse as h . This experiment shows that at both 0° and 45° the direct method produces similar, but rotated results, whereas the first-order method result is much more diffuse for the 45° direction than for 0° . The difference is especially apparent close to the impulse: at 45° the first-order method result is not nearly as focused as the direct method result. See Fig. 6.1. In contour filtering applications, accurate results are especially important nearby the impulse, where most of the filtering action occurs. In computer vision, the damaging effect of the discrete grid has often been observed, where vertical and horizontal edges are much more easily enhanced than oblique ones [128, p. 45]. The direct method overcomes this anisotropy in the context of the direction process.

⁴Since the particle simulations would be done only once, excessive computation times may be tolerable.

⁵Twice convolved with a $[1\ 1]$ box filter in the x - and y -directions.

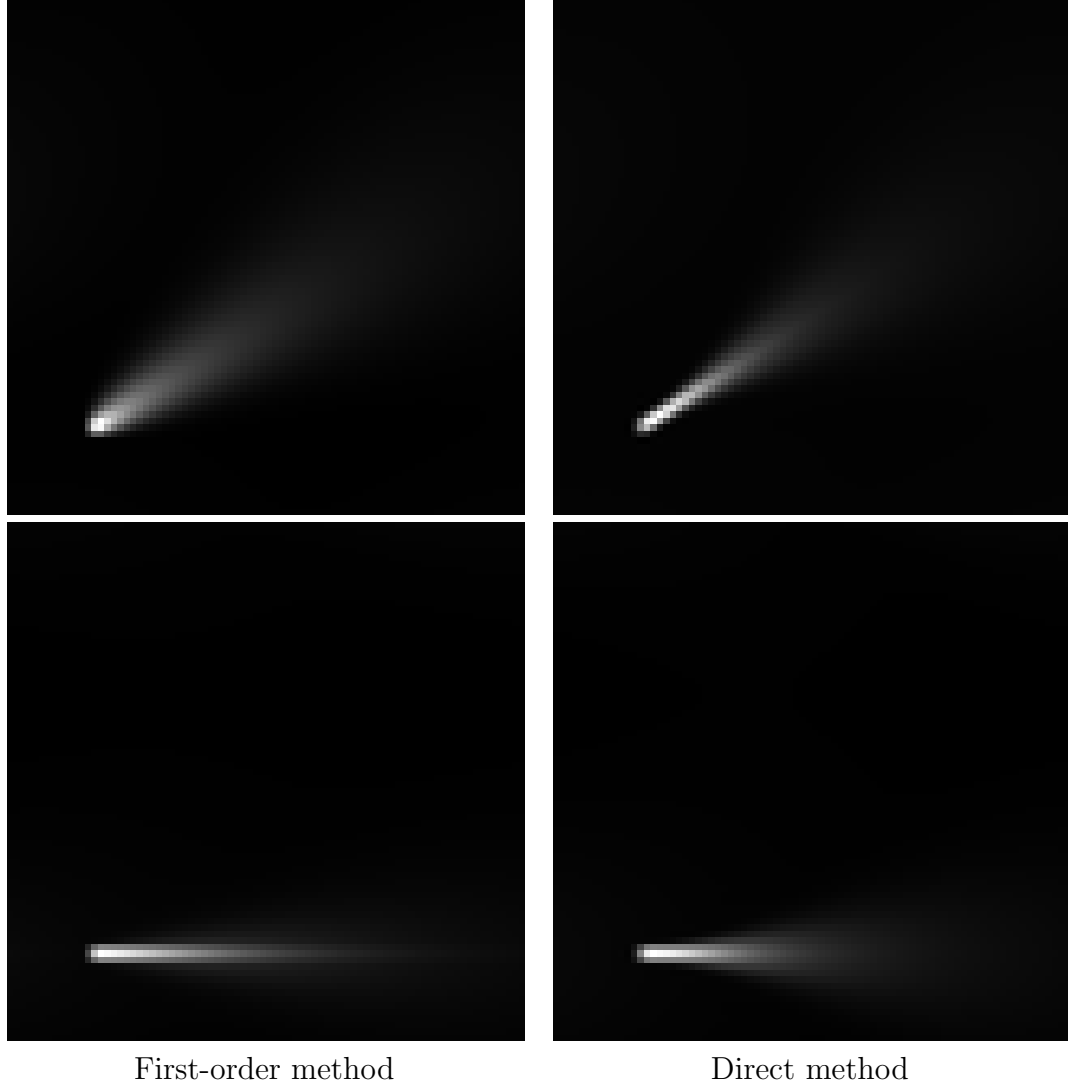


Figure 6.1: Isotropy comparison. The direction process forward PDE was solved for a slightly blurred impulse on a $80 \times 80 \times 44$ grid, with parameters $\sigma_\kappa = 1/24$, $\alpha = 1/100$, and at discrete directions 4 (top) and 0 (bottom). The spatial discretization was set to $\Delta x = \Delta y = 1$ here and elsewhere in this thesis, unless stated otherwise. Depicted is the integral over θ . The direct method produces a more focused result at discrete direction 4 than does the first-order method; the first-order method also has insufficient blurriness far to the right of the horizontal impulse (lower left), unlike the direct method result (lower right).

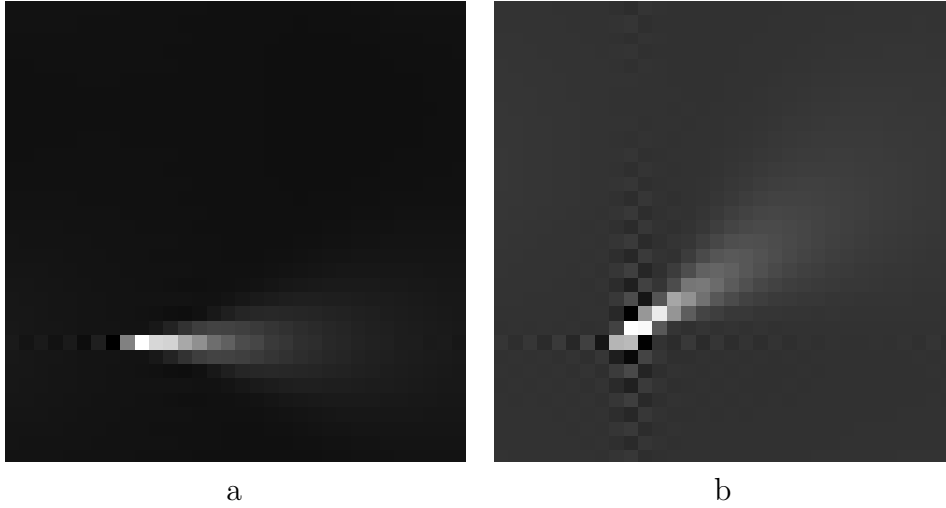


Figure 6.2: The discrete impulse is the sampling of a (periodic) sinc function, giving rise to the oscillations observed in these solutions to the direction process forward PDE with the direct method. The initial distribution h is an (unblurred) impulse at $(8, 8, 0)$ in (a) and $(8, 8, 4)$ in (b). Parameters are $\sigma_\kappa = 0.1, \alpha = 1/100$ on a $32 \times 32 \times 44$ grid.

For the second experiment, we show the oscillations that occur in the direct solution when the (sampling of) h is a discrete impulse (Fig. 6.2). This phenomenon motivates the slight Gaussian pre-blurring that we used in Fig. 6.1 and elsewhere in this thesis.

The next experiment was designed to illustrate the accuracy of the direct solution compared to finite differencing. A fixed stimulus h was chosen, e.g. a Gaussian, and the discretization was varied on a regular schedule to see the improvement in solution accuracy with successive doublings of spatial resolution. See Fig. 6.3. The relative residual in the conjugate gradient iterations (see §6.2, except set $c = 0$) for the first-order method were held to less than 10^{-10} , and up to 10,000 iterations were allowed (but fewer iterations were generally needed to achieve the desired relative residual). Fig. 6.4 shows the convergence of the first-order method, and reveals that larger resolutions converge much more slowly. The computational burden of the iterative first-order method was many-fold longer in the larger resolutions, taking on

the order of a day on a 600 MHz Pentium II. The direct method only took half a minute.

To put the direct and first-order methods into clearer contrast, we re-plotted columns (b) and (c) of Fig. 6.3 to show, for each method, the *rate* at which the solution is changing (which we assume to mean improving) with increasing resolution. To save storage space, computation, and, most importantly, achieve a solution accuracy as close to machine precision as desired, it is important that the solution improves quickly as a function of increasing resolution. For each method, we measured the relative change in the solution over the common 8×8 subgrid from scale k to scale $k + 1$ as a function of scale. Let $h_{8 \times 8}(2^k)$ be the solution obtained on the 8×8 subgrid at resolution $2^k \times 2^k$. The relative change in solution is

$$\frac{||h_{8 \times 8}(2^{k+1}) - h_{8 \times 8}(2^k)||}{||h_{8 \times 8}(2^{k+1})||}.$$

See Fig. 6.5. For the first-order method, we observe that the relative change in solution is greater than 10% for the 8×8 to 16×16 resolution change. This relative change in solution decreases very slightly, so that after four doublings of resolution, the relative change is still several percent. For the direct method, the results are dramatic: at first the relative change in solution is about the same as for the first-order method. But by the 64×64 to 128×128 resolution change, the relative change is on the order of machine precision (about 10^{-15}). The small improvement for the next increase in resolution further illustrates that the machine precision limit has been achieved. The direct and first-order methods were implemented in Python with the numerical Python library and C extensions under GNU/Linux.

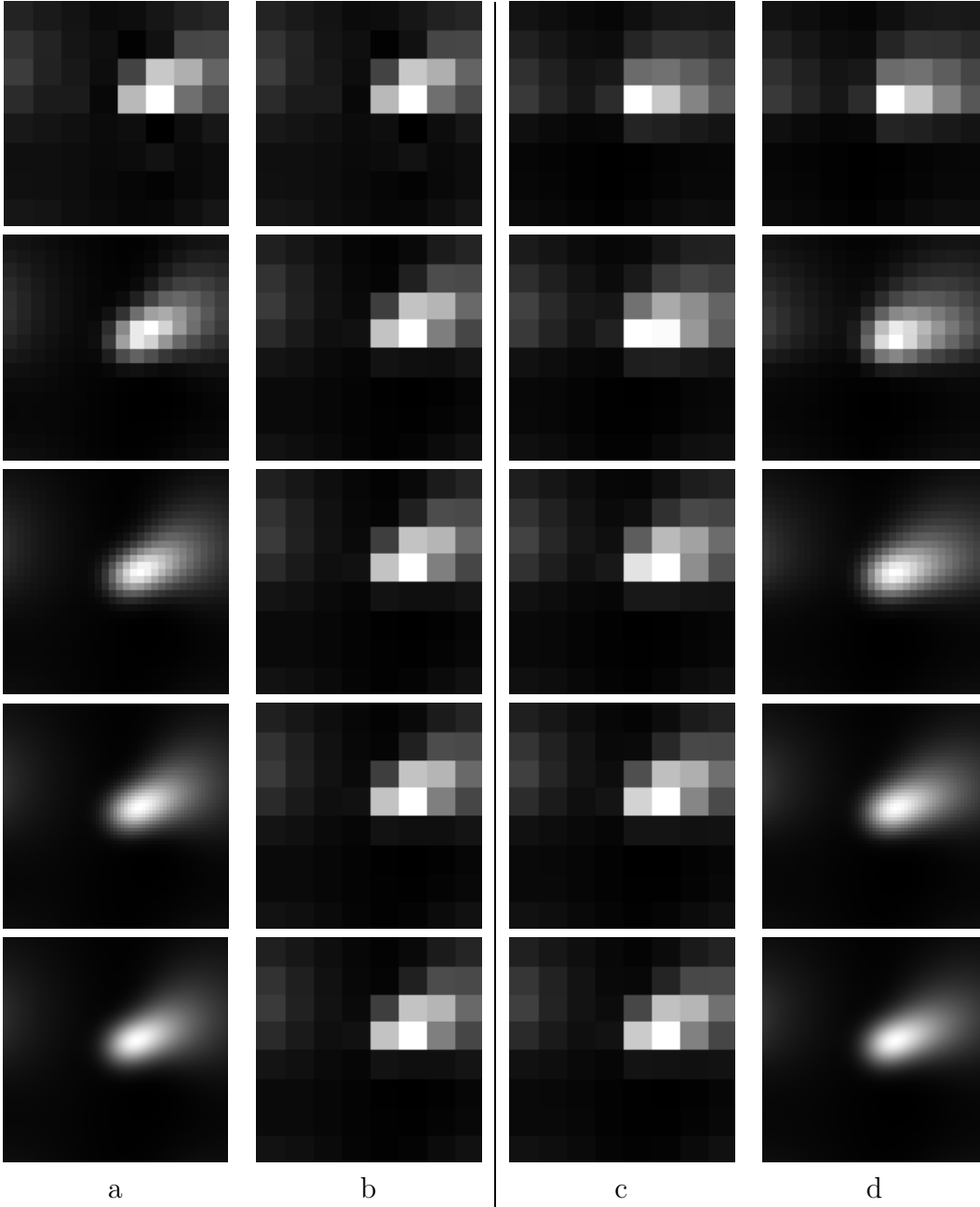


Figure 6.3: Comparison of solution improvement with increased spatial resolution. The direction process forward PDE was solved on $[0, 1]^2 \times [0, 2\pi]$ with periodic boundary conditions for input b a Gaussian centered at $(1/2, 1/2, 3/8)$ with standard deviation $1/16$ in all three dimensions. Parameter values are $\sigma_\kappa = \alpha = 1$. In column (a) are shown the integrals over θ for the direct method solution, for discretization $n \times n \times 16$, where $n = 8, 16, 32, 64, 128$ (top to bottom) and $\Delta x = \Delta y = 1/n$. To facilitate comparison, column (b) shows the solution in column (a) subsampled to an 8×8 grid. Observe the slight change from the 8×8 to the 16×16 resolution, with no visible improvement thereafter (also see Fig. 6.5). In columns (d) and (c) are the corresponding results for the first-order method.

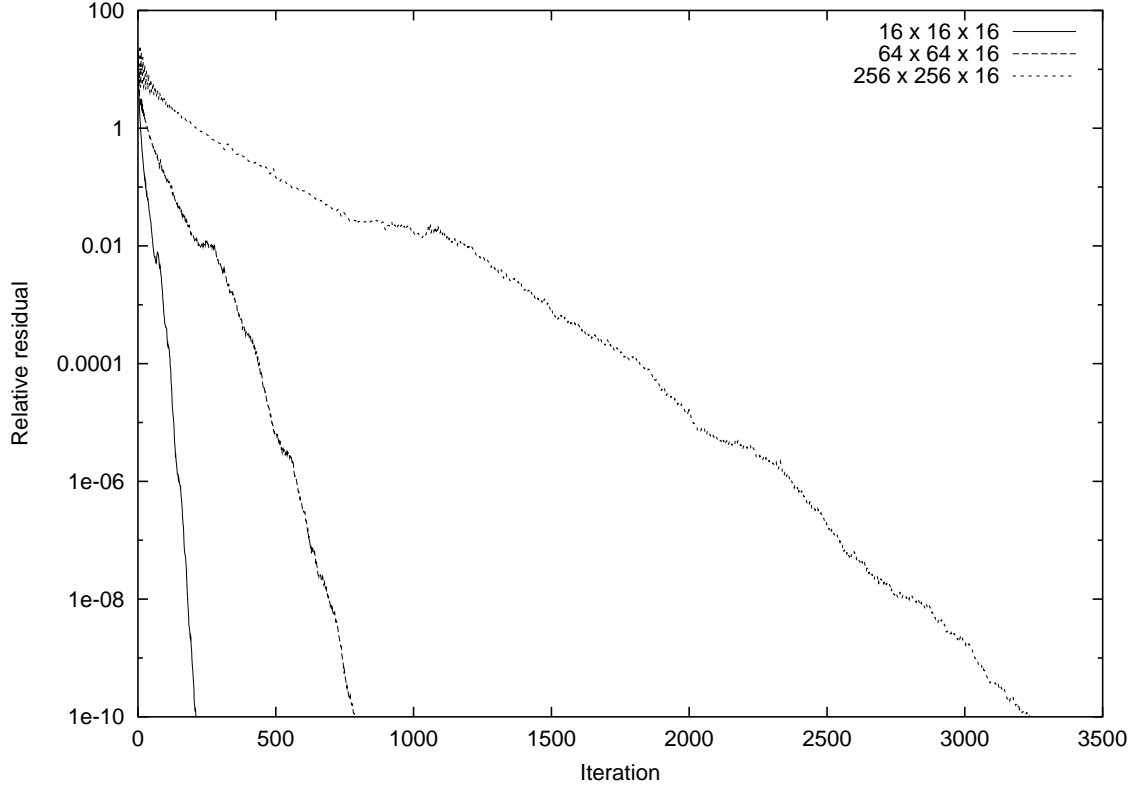


Figure 6.4: Decrease in relative residual (for the first-order method of computing the direction process forward PDE solution) with iterations of conjugate gradient method. Each plot is for the same Gaussian input (see text) sampled at the resolutions shown.

6.1.2 Curvature Process

To compute the action f of the Green's operator G for the curvature process on a vector h , we recall the curvature process generator (2.7) to set up the *curvature process forward PDE*

$$\alpha f - \frac{\sigma^2}{2} \frac{\partial^2 f}{\partial \kappa^2} + \cos \theta \frac{\partial f}{\partial x} + \sin \theta \frac{\partial f}{\partial y} + \kappa \frac{\partial f}{\partial \theta} = h, \quad (6.5)$$

where $\sigma = \sigma_{\kappa} > 0$. The action of the Green's operator G is then defined as our numerical solution to (6.5). To balance performance and simplicity, we use a combination of finite differencing and basis projection. To keep the desired rotational isotropy and solution focus, we project f with respect to *only* the x -, y -, and θ -

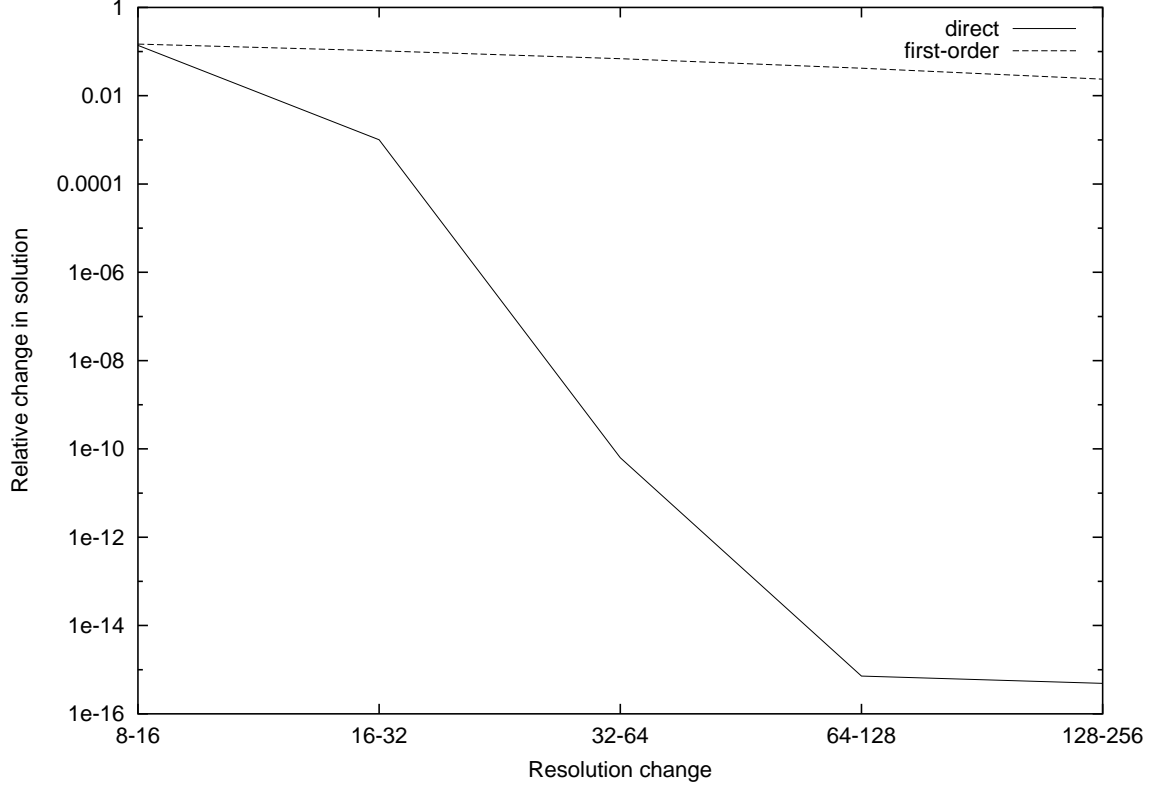


Figure 6.5: Rate of improvement in solution with increase in resolution. See text.

dimensions on the same complex exponential basis that we used for the direction process to obtain $\hat{f}_{jkl}(\kappa) := \langle f(\cdot, \cdot, \cdot, \kappa), \phi_{jkl} \rangle_{\mathbb{R}^2 \times \mathbb{S}}$ for frequencies j , k , and l . The κ -dimension is discretized into M uniform samples using finite differencing, as follows. We first write $\hat{f}_{jkl;m'} = \hat{f}_{jkl}(m'\Delta\kappa)$, where $\kappa = m'\Delta\kappa$ for integer m' and stepsize $\Delta\kappa > 0$. The input h is discretized similarly. Then we approximate $\frac{\partial^2 f}{\partial \kappa^2}$ with $(g_{m'-1} - 2g_{m'} + g_{m'+1})/(\Delta\kappa)^2$, for $g_{m'} := f(x, y, \theta, m'\Delta\kappa)$.

Taking the (x, y, θ) -projection of (6.5) onto the complex exponential basis, we can reuse the direction process results, noting that $\langle \frac{\partial f}{\partial \theta}, \phi_{jkl} \rangle_{\mathbb{R}^2 \times \mathbb{S}} = \mathbb{I} \hat{f}_{jkl;m'}$. We assume the same bandlimitedness (with respect to (x, y, θ)) of h as in the direction process, and so are led to solve for the lowest J, K, L frequencies in (x, y, θ) of f in terms of those of h . Combined with the finite difference approximation in κ , we now need to

solve

$$\begin{aligned} & \left(\alpha + \frac{\sigma^2}{(\Delta\kappa)^2} + \mathfrak{I}lm'\Delta\kappa \right) \hat{f}_{j,k,l;m'} - \frac{\sigma^2}{2(\Delta\kappa)^2} \left(\hat{f}_{j,k,l;m'-1} + \hat{f}_{j,k,l;m'+1} \right) \\ & + \pi \left(\frac{\mathfrak{I}j}{x_{\max}} + \frac{k}{y_{\max}} \right) \hat{f}_{j,k,l-1;m'} + \pi \left(\frac{\mathfrak{I}j}{x_{\max}} - \frac{k}{y_{\max}} \right) \hat{f}_{j,k,l+1;m'} = \hat{h}_{j,k,l;m'}. \end{aligned} \quad (6.6)$$

The same periodic boundary condition in l is used as in the direction process case.

We use zero boundary conditions in m' to put an upper bound on curvature.

Observe that for each j and k , equation (6.6) is an $(L \cdot M) \times (L \cdot M)$ sparse linear system in l and m' . Although the computationally advantageous tridiagonality of the directional process case has been lost, we note that we are often interested in small L and M (the numbers of discrete directions and curvatures, respectively). The overall complexity of this system is $O((JKLM) \log(JKLM)) + O(JK(LM)^3)$, using LU decomposition to solve the linear system.

6.2 Applying the Biased Green's Operator

In this section we present a method for computing the action f of the biased Green's operator G_c on a vector h . Equivalently, we must solve the linear system

$$-(Q + C)f = h \quad (6.7)$$

where $Q = \alpha I - L$ and C is a (possibly non-constant) diagonal operator with $c = (c_i)$ along the diagonal. See §3.3 for how this equation arises in the Feynman-Kac formula. Because of Khas'sminskii's condition (Proposition 3), we know that there is a unique solution to this equation as long as c is sufficiently small.

We first assume that we are given a method for solving $-Qf = h$, i.e.:

$$f = Gh.$$

Such methods were just provided in §6.1.1 and §6.1.2 for the direction and curvature processes, respectively. Observe that G is a discretization of an integral operator, as it is the inverse of a discretization $\alpha I - L$ of differential operator $\alpha - \mathcal{L}$. Loosely, recall that, for the direction process, G represents a kind of blurring of a directed edge map (function of (x, y, θ)) in the tangent direction θ .

Now, instead of solving the (6.7) as (a discretization of) a PDE, we will convert it into (a discretization of) an integral equation by multiplying (6.7) by G on the left:

$$A_1 f := G(-Q - C)f = (I - GC)f = Gh =: b_1.$$

This procedure is known as preconditioning a linear system. Since we multiplied $-Q - C$ by G on the left, we shall say that $A_1 u = b_1$ is preconditioned on the left. One can instead make the substitution $f = Gv$ and solve the linear system:

$$A_2 v := (I - CG)v = (-Q - C)Gv = (-Q - C)f = h =: b_2.$$

Since G appears to the right of $-Q - C$, we shall say that $A_1 u = b_1$ is preconditioned on the right. The idea behind preconditioning is that we turn a system that may be badly behaved (such as a differential equation) into one which is better behaved, such as an integral equation.

To solve $Au = b$, where $A = A_i, b = b_i, i = 1, 2$, we will invoke the conjugate gradient method [56]. Unfortunately, this method requires a symmetric and positive definite linear system, but Q and therefore A is not symmetric. One technique for

circumventing the asymmetry of A is to multiply (6.7) on the left by its transpose A^* , forming the *normal equations*

$$A^*Au = A^*b, \quad (6.8)$$

where A^*A is symmetric because $(A^*A)^* = A^*(A^*)^* = A^*A$ and positive definite because $x^*A^*Ax = (Ax)^*(Ax) = \|Ax\|^2 > 0$ for nonsingular A unless $x = 0$. Now, we can apply the (iterative) conjugate gradient method directly to (6.8) in order to solve for u in the sense that the relative residual $\|A^*Au - A^*b\|/\|A^*b\|$ is reduced to less than some tolerance level. We can then apply the above definitions to get f itself.

6.3 Solving for the Biased CIRF Approximation of the Posterior

In §5.7 we derived coupled nonlinear systems for filtering. Because of the presence of blur in a number of applications, we elected to solve the nonlinear system in equations (5.32) and (5.33). Left multiplying these by Q^{-1} and Q^{*-1} respectively, we get the following system:

$$f - \epsilon GCf + \nu' = 0 \quad (6.9)$$

$$b - \epsilon G^*Cb + \mu' = 0, \quad (6.10)$$

where $C := \text{diag}(a + 2\bar{\mathcal{N}}A(f \odot b))$, $\nu' = G\nu$, and $\mu' = G^*\mu$. To solve this nonlinear system, we adopted a continuation strategy in ϵ . Specifically, we began with ϵ small, solved the system, and then used that solution to get an initial condition for the

system with ϵ slightly larger. We proceeded in this fashion until we reached the value of ϵ required. To solve this system for any such fixed ϵ , we used Newton's method [56], where often only a few iterations are required because of the Newton method's powerful local convergence properties. To bootstrap this continuation process with a starting initial condition (corresponding to the smallest value of ϵ), we set $C = \text{diag}(a)$ (i.e., neglecting the nonlinearity) and then solved the resulting decoupled equations, which evidently are the biased CIRF Green's operator equations that we saw in the previous section. We now apply our filters to images.

Chapter 7

Results

In this chapter we illustrate the performance of CIRF filters. We begin by applying direction process-based Volterra filters to some synthetic images and later to a few natural images. The Volterra filter is also briefly compared to the biased CIRF approximation filter, where the latter shows more powerful noise cleaning action. We conclude with some examples of the curvature process-based CIRF filter.

7.1 Contour Filtering in $\mathbb{R}^2 \times \mathbb{S}$ with Direction

The first test of the CIRF framework was with the MMSE Volterra filters using the direction process in $\mathbb{R}^2 \times \mathbb{S}$. With the direct method for computing the action of G and G^* from the last chapter, applying these filters requires substitution into the Volterra formulas in Result 1. We selected parameters as follows. Because the density parameter η is an overall multiplicative constant, it amounts to a threshold setting. We developed a user interface to select the parameters ϵ, ζ , the threshold, and the order of the filter. By inspection of the Volterra filtering formulas, one can do most of the computation before parameter selection. Then, for each parameter setting, only a few fields need to be weighted and added; this means that parameters

can be chosen rapidly; typically, selecting parameters took less than a minute. In most examples we let ζ be small (usually 0) and let ϵ be about 10, although the exact setting of ϵ did not appear to have a great effect. In the end the user interface was most useful for setting the threshold and comparing the different orders of filter.

Because of the computational efficiency of the Volterra filters, we were able to test them on a number of examples. To assess the robustness of the Volterra filter in the presence of noise, we evaluated responses to images of a horizontal straight line corrupted by additive Gaussian noise of variance σ_N^2 . We characterized the degree of noisiness using the peak signal-to-noise ratio, which, for an image \mathbf{I} taking values in the range \mathbf{I}_{\min} to \mathbf{I}_{\max} , is

$$\text{Peak SNR} = 20 \log_{10} \frac{\mathbf{I}_{\max} - \mathbf{I}_{\min}}{\sigma_N},$$

expressed in decibels. To map from the planar image to a field on discrete $\mathbb{R}^2 \times \mathbb{S}$, we simply copied the image to each direction, so that our measurements $m = m(x, y, \theta)$ were constant in θ . We were able to select this mapping because the images were purely contours embedded in noise. More generally an oriented local edge or line operator would be applied, as we shall see when we consider natural images.

To avoid the aliasing effect illustrated in Fig. 6.2, for all experiments in this chapter we convolved the input with a Gaussian having a standard deviation of 1 pixel in all dimensions before applying a CIRF filter. The results of applying the linear, quadratic, and cubic Volterra filters to the noisy line image are shown in Figs. 7.1 and 7.2, where the display is produced by taking the maximum response over all directions θ for each position (x, y) . Observe that while the filter responses are strongest along the line, there is some response elsewhere, especially for the linear filters. This extra response can be thought of as weaker contour hypotheses, and

conveys how a field model can parsimoniously express the infinite space of contour explanations. We also observe that as the order of the filter is increased to quadratic and then cubic, the response becomes increasingly focused on the desired line. To quantify these results, we placed the responses as vectors in very high-dimensional spaces. For these $64 \times 64 \times 64$ fields, the dimension of this space is 262,144. Then, each response was then centered by subtracting its average value from itself. For each kind of filter order (linear, quadratic, and cubic), we measured the angle between the (high-dimensional) vector \tilde{u}_1 representing the noiseless (centered) response and the vector \tilde{u}_2 representing the (centered) response under a particular level of noise. The angle (in degrees) between centered responses \tilde{u}_1 and \tilde{u}_2 is defined as:

$$\frac{180^\circ}{\pi} \cos^{-1} \frac{\langle \tilde{u}_1, \tilde{u}_2 \rangle}{\|\tilde{u}_1\| \|\tilde{u}_2\|},$$

where $\langle \tilde{u}_1, \tilde{u}_2 \rangle$ is the Euclidean inner product between \tilde{u}_1 and \tilde{u}_2 , and $\|u\| = \sqrt{\langle u, u \rangle}$. Observe the quantitative improvement of the quadratic and cubic nonlinear filters compared to the linear result (Fig. 7.3); this nonlinear advantage appears even greater as σ_κ is made smaller, i.e., when the contours are assumed to be straighter.

To study the effect that the parameters of the direction process have on our filtering results, we created an image containing curves of different curvatures and lines with gaps, and then applied the direction process-based MMSE CIRF cubic Volterra filter at a number of parameter settings (Fig. 7.4). The number of directions used in each test was 16, 16, 32, and 64, for σ_κ settings of 1, 0.33, 0.2, and 0.1, respectively; using more directions has a negligible visible effect. The image at the top was input to the filter by copying the image over all directions θ (black is 0, and white is 1); no noise was added, although discretization artifacts are evident in the semi-circular contour regions. The brightness of the filter outputs (bottom four rows) was obtained

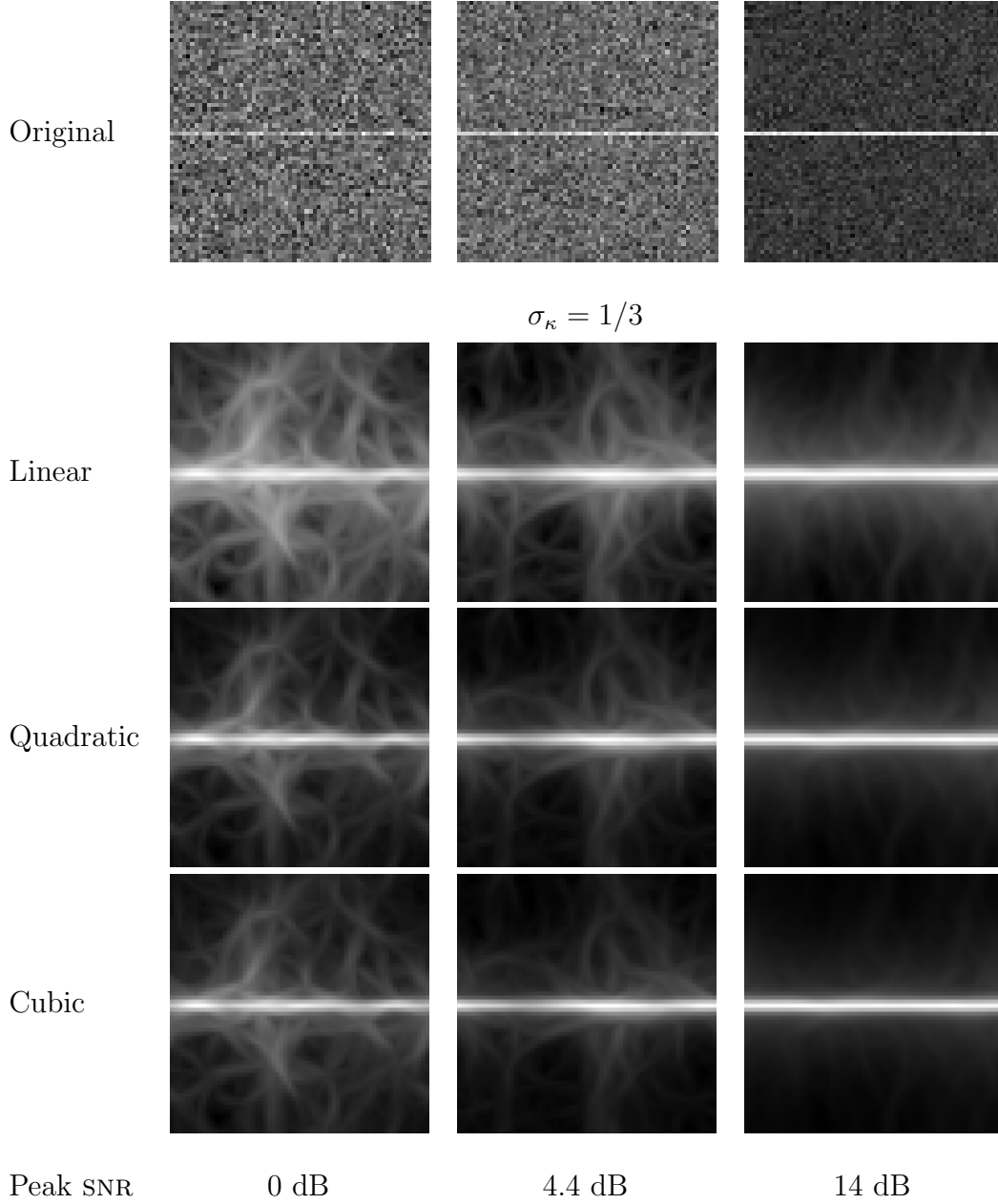


Figure 7.1: Enhancing a line in the presence of noise (see text for details). The original images are in the top row, and the corresponding direction process-based MMSE CIRF Volterra filter responses are as indicated in the rows below. Parameters are $\epsilon = 10$, $\zeta = 0$, $\sigma_\kappa = 1/3$, $\lambda = 100$, and 64 directions. Continued in Fig. 7.2. For a quantitative comparison, see Fig. 7.3.

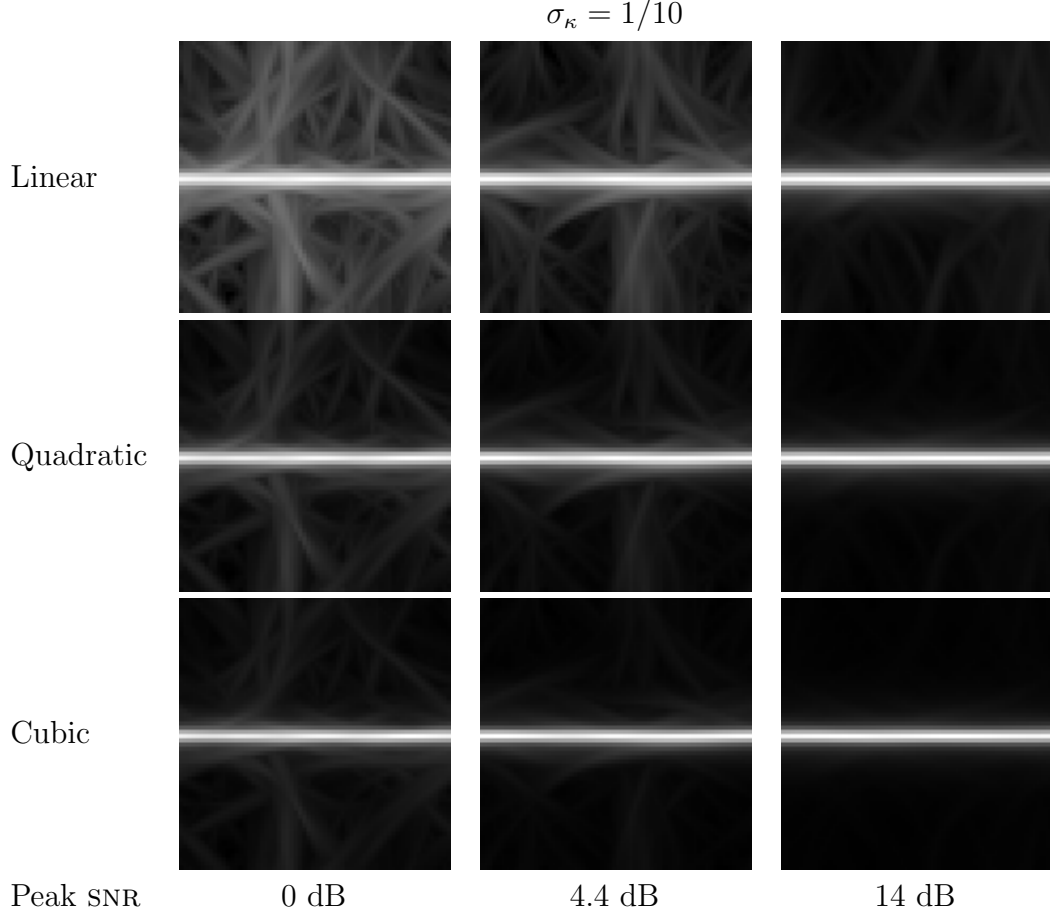


Figure 7.2: Enhancing a line in the presence of noise (continued from Fig. 7.1). Parameters are $\epsilon = 10, \zeta = 0, \sigma_\kappa = 1/10, \lambda = 100$, and 64 directions. Observe the much-reduced response away from the curve compared to the case $\sigma_\kappa = 1/3$ in Fig. 7.1, especially in the quadratic and cubic Volterra responses. For a quantitative comparison, see Fig. 7.3.

from the maximum response over all directions θ at each position (x, y) . Values of σ_κ that we have often used in experimenting with the Volterra filters correspond to the bottom three rows. For $\lambda = 100$ (a typical value in our work), we observe that the gaps are partially filled at $\sigma_\kappa = 0.33$, and are more strongly completed as σ_κ is decreased to 0.1. However, the small half-circle in the center is attenuated at small σ_κ , as large curvatures in the direction process become more improbable. Because the filter output is not sensitive to variations in λ when λ is large (around 100, say), we used a logarithmic scale in λ . For $\lambda = 10$, slightly greater attenuation in the gaps

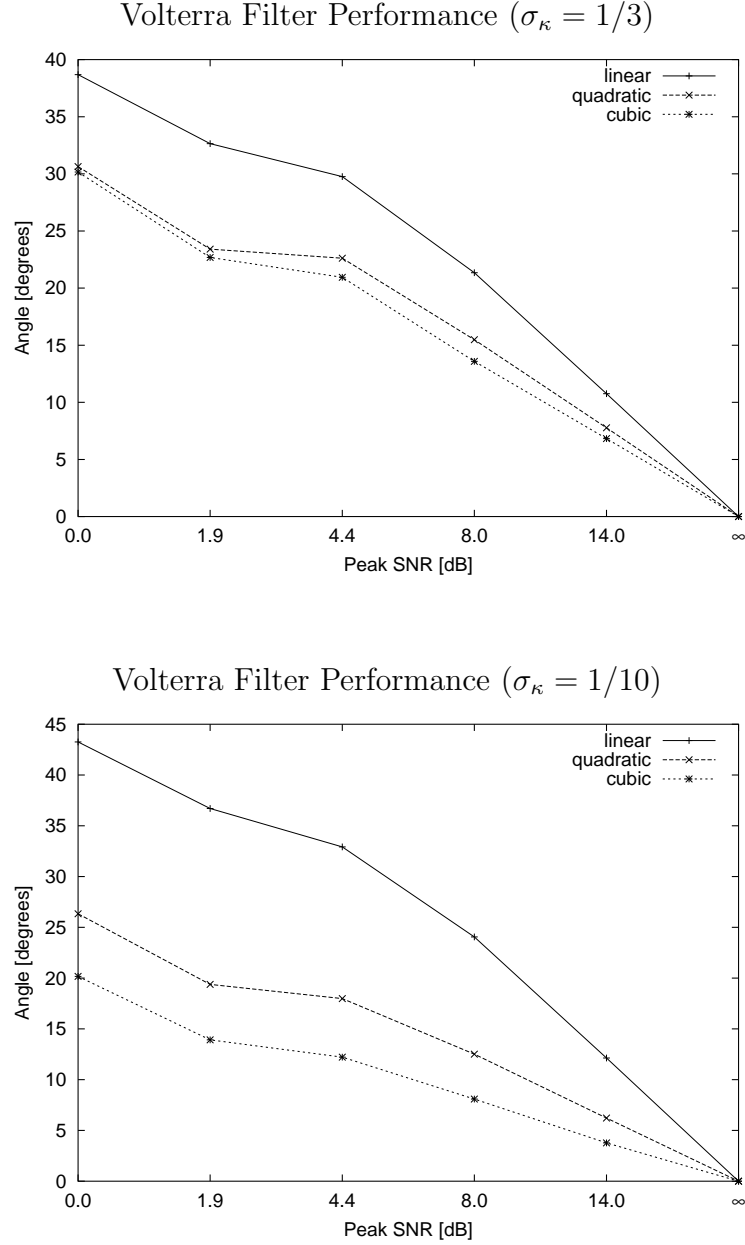


Figure 7.3: Noise performance of direction process-based MMSE CIRF Volterra filters. Plotted is the angle between the centered filter responses to the noisy and noise-free images shown in Fig. 7.1 (top plot above) and Fig. 7.2 (bottom plot above); see text for details.

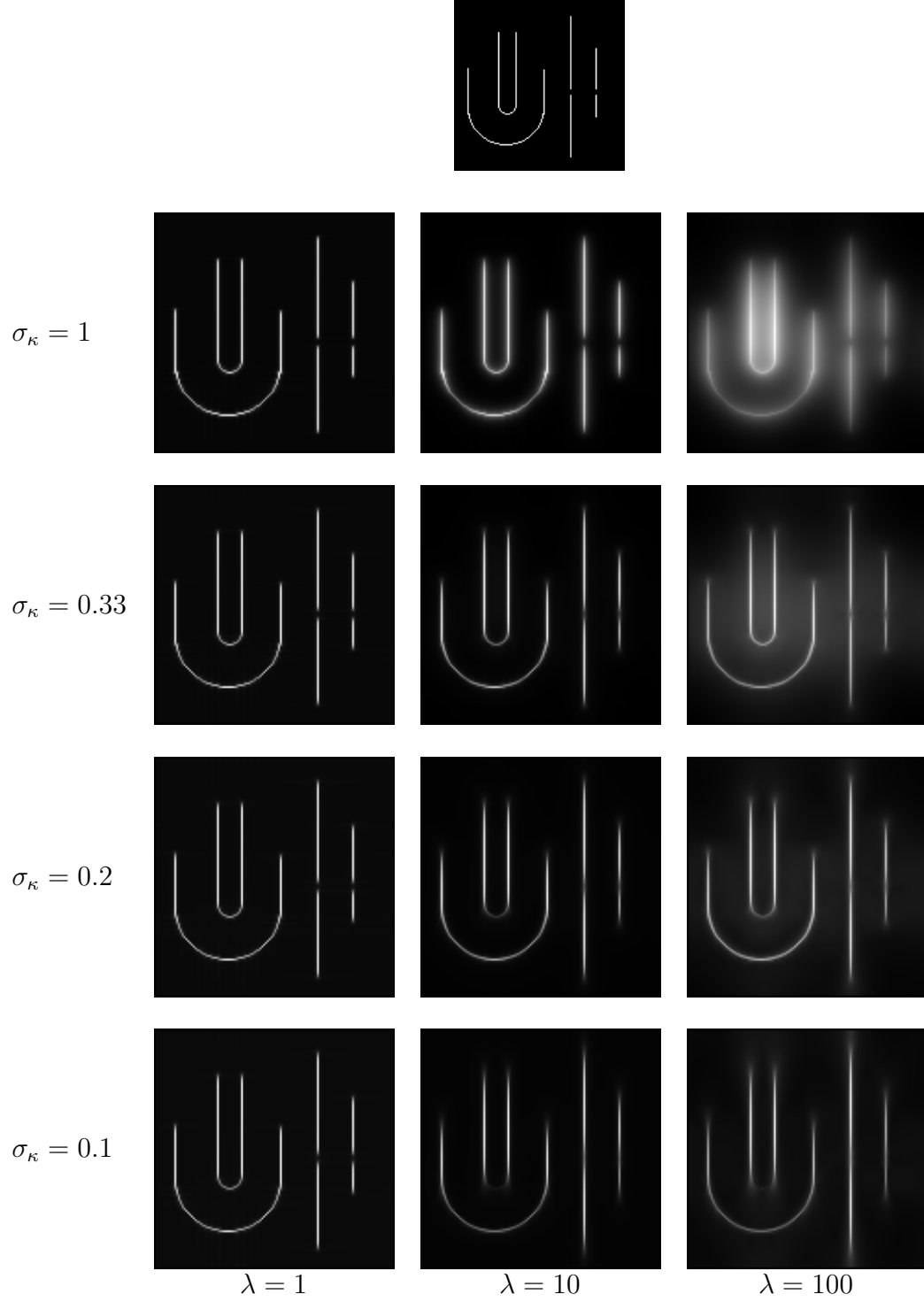


Figure 7.4: The effect of varying the direction process parameters σ_κ (bottom four rows, top to bottom) and λ (left to right) on the cubic MMSE CIRF Volterra filter output for the input at the top. See text for explanation.

and at the small semi-circle can be seen. When $\lambda = 1$, the filter is effectively “turned off”. For comparison, we also include the results for $\sigma_\kappa = 1$, which corresponds to curves with little smoothness (they approach planar Brownian motion-like behavior). This experiment emphasizes the tradeoff between resiliency to noise (as represented by the gaps) and performance at high curvatures.

We then tested the filter’s response on both open and closed curves (Fig. 7.5). Despite the crossings of the curves, the noise cleaning action is sufficiently good that simple thresholding is sufficient. Observe carefully the response is well resolved at the contour crossings; this happens even though we have not included the blur inherent in the loss of local direction in the original 2-dimensional image: by copying the input image over θ to produce the input, we have effectively produced the maximum possible blur in θ .

To test the behaviour of the filter at crossings more strenuously, we created an image with several nearby crossings (Fig. 7.6). Such crossings pose a problem for the isotropic smoothing often applied to cope with noise: the nearby junctions will merge with a sufficiently large blur kernel. The Volterra filter does perform smoothing of a sort, but primarily along contours, and therefore the preservation of junctions is possible. This crossing-preservation behaviour is exactly what is observed in the cubic Volterra result (Fig. 7.6, top, second from right); in addition, quite powerful orientation-selectivity is apparent (fixed θ slices of the response are shown in row (d) of Fig. 7.6).

We next applied the Volterra filters to two natural images. Instead of simply copying the image over directions to form the input, we obtained the filter input m from the output of the logical/linear operator for negative-contrast lines (with parameter degree $d = 0.1$ in Iverson’s implementation,¹ and other values at their

¹Code and settings available at <http://www.ai.sri.com/~leei/loglin.html>.

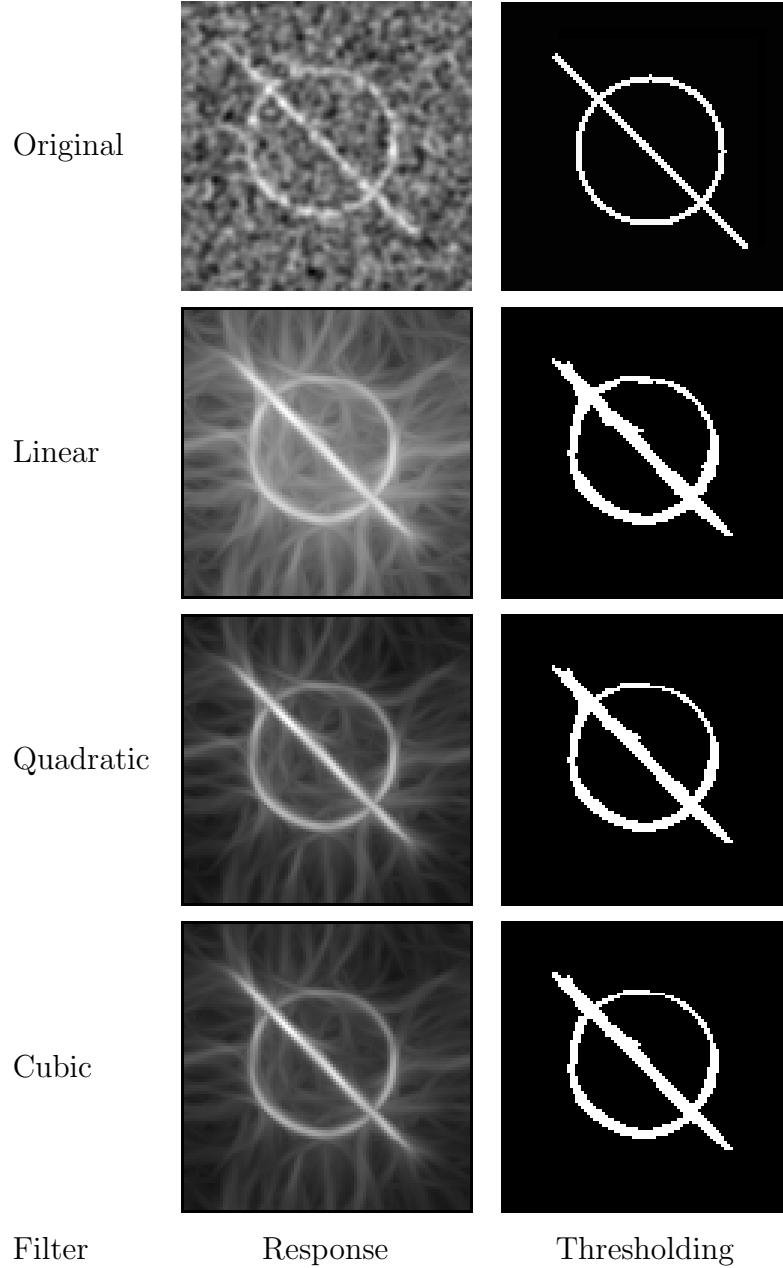


Figure 7.5: Filtering crossing contours with direction process-based Volterra filters ($\epsilon = 10.5, \zeta = 0, \sigma_\kappa = 1/5, \lambda = 100$, and 32 directions). Image before adding white noise is at the top right; the noisy image (peak SNR=4.2 dB) at the top left was the input (copied over all directions θ). Observe that both curves are preserved in the filter responses (bottom three rows, left) and their thresholding (bottom three rows, right).

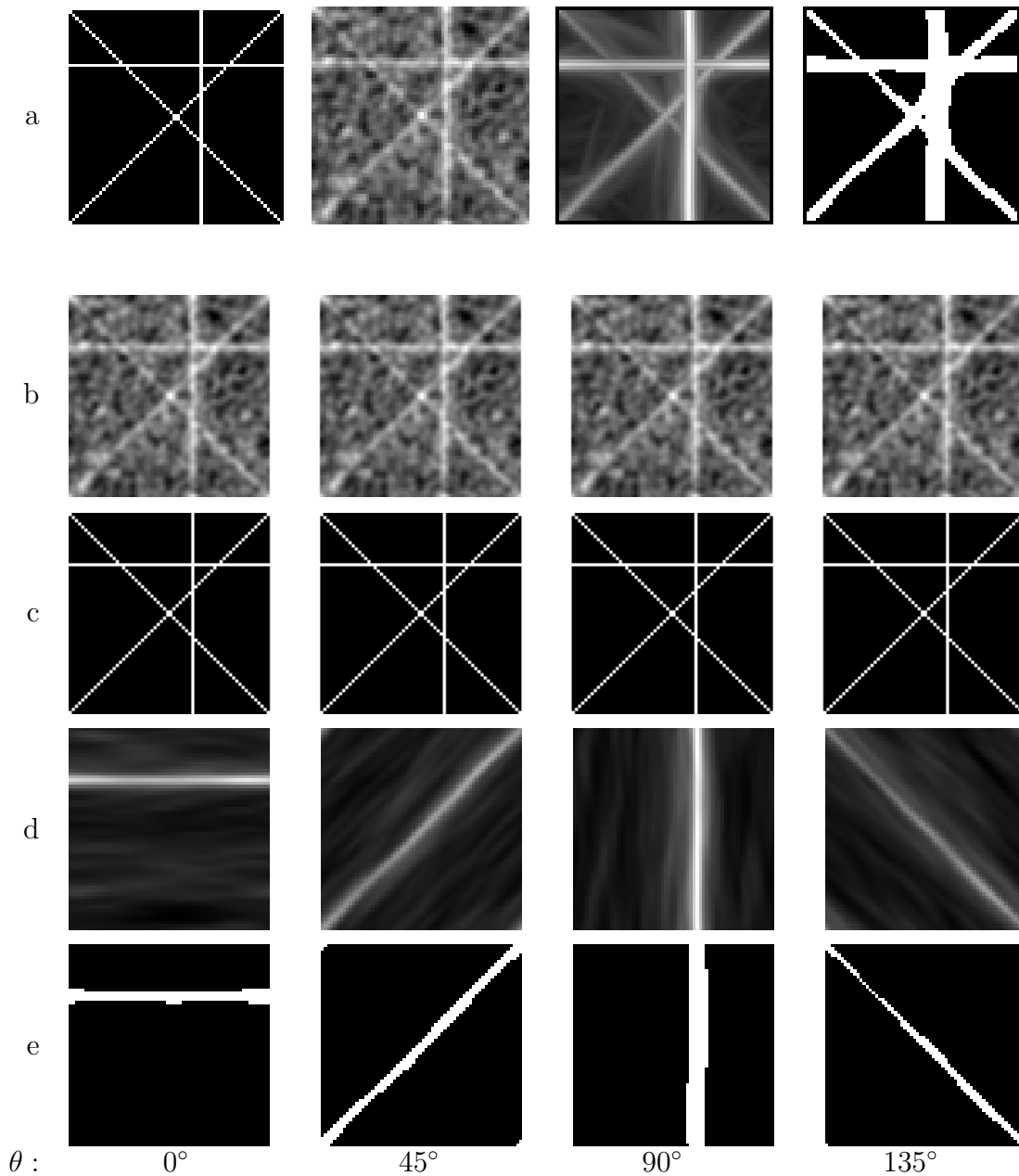


Figure 7.6: Crossing lines are teased apart with the cubic Volterra filter. From left to right in row (a): original image, image corrupted with additive white Gaussian noise (peak SNR=8 dB), cubic Volterra filter response ($\epsilon = 10.5, \zeta = 0, \sigma_\kappa = 1/10, \lambda = 100$, and 64 directions), and thresholded response. Noisy image was copied over all directions to form filter input (row (b); for comparison we include the corresponding noise-free "stack" of direction-specific images in row (c)), then filtered with the high-noise MMSE CIRF cubic Volterra filter (row (d)), and thresholded (row (e)). Observe how the initially strong responses at the inappropriate directions are chiseled away by the action of the filter, despite the presence of noise.

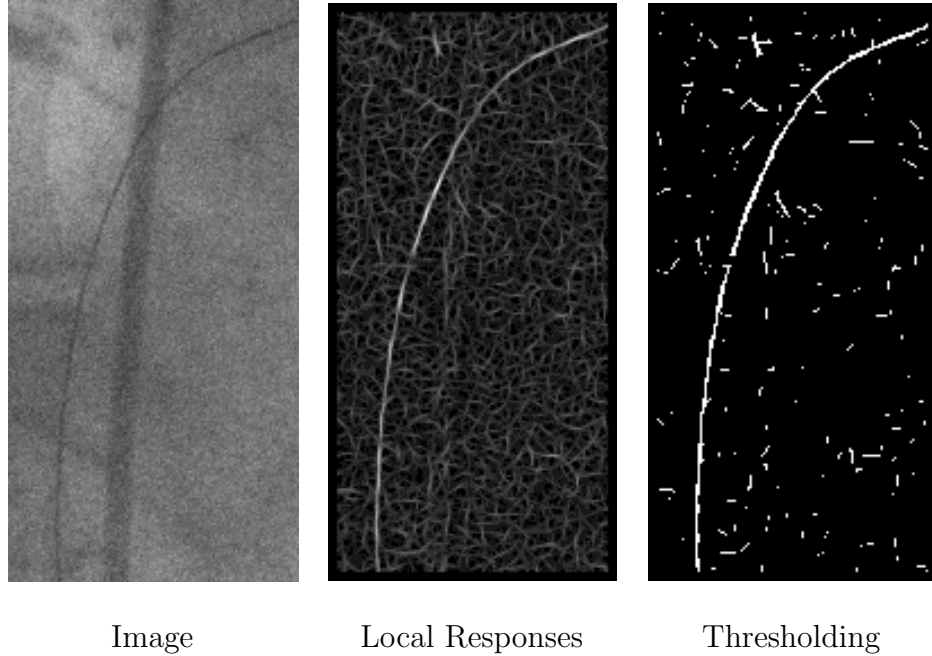


Figure 7.7: Fluoroscopic image of guide wire (left), logical/linear negative-contrast line responses (center) and their thresholding (right). See Fig. 7.8 for Volterra filtering results.

default). The first example is of a guide wire imaged fluoroscopically during surgery (Fig. 7.7). Finding the wire in noise is particularly important, because surgery can take hours, all the while exposing the patient to radiation. Reducing such radiation exposures unfortunately also increases the noise level and makes *local* filtering (e.g., edge and line operators) perform more poorly. With the MMSE CIRF Volterra filters, we can readily enhance the guide wire contour (Fig. 7.8).

Another example is the ship's wake shown in Fig. 7.9 (top left). For this demonstration, we work with the magnified section shown in Fig. 7.9 (top right), and then obtain logical/linear negative-contrast line responses (with degree 0.1) and their thresholding (second row). Responses to high-noise Volterra filters (Result 1) for the direction process-based CIRF ($\epsilon = 10.5, \zeta = 0, \sigma_\kappa = 1/15, \lambda = 100$, and 96 directions) is shown in the bottom three rows of Fig. 7.9. To produce these two dimensional images from the actual discrete (x, y, θ) -space responses, the maximum over θ was

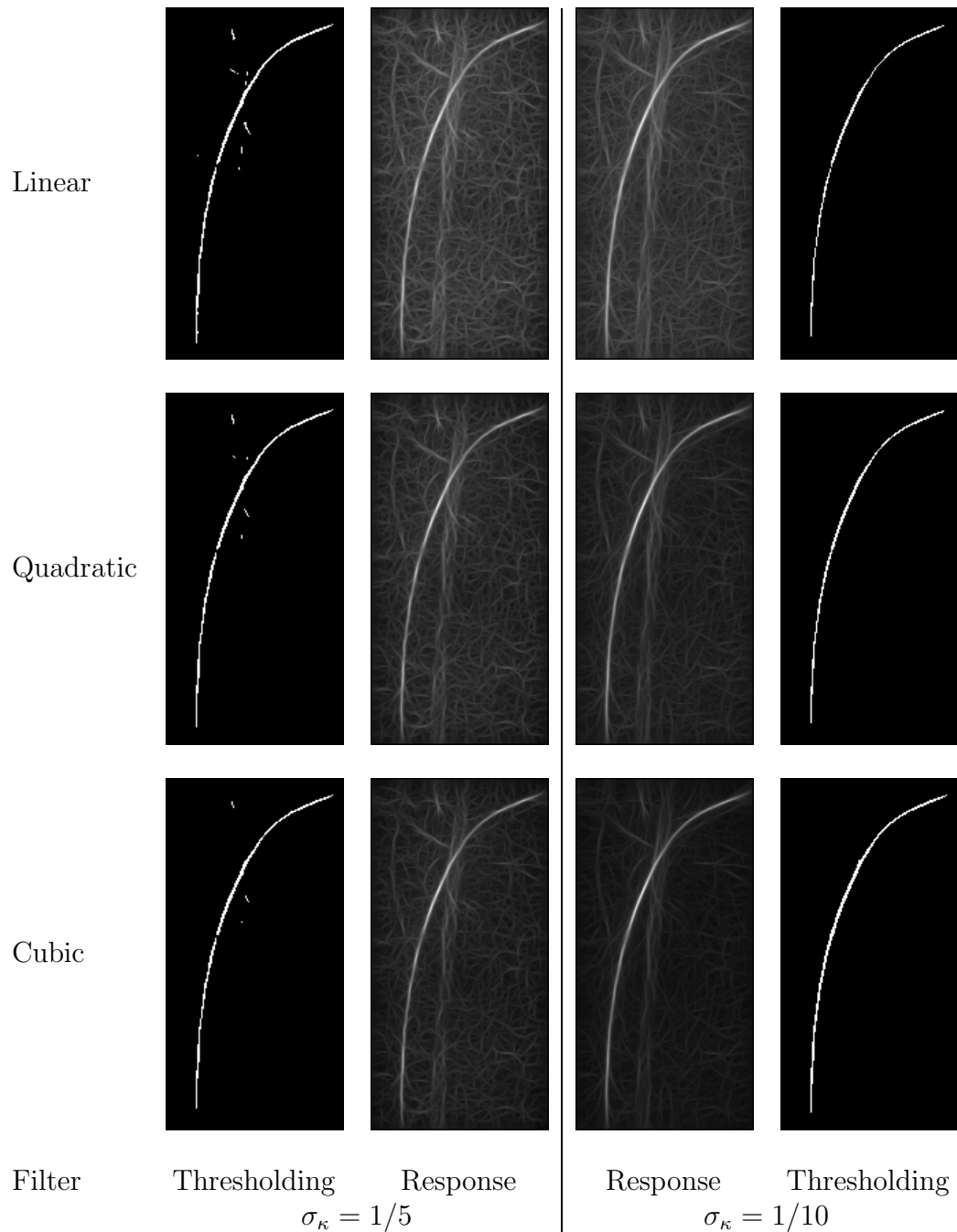


Figure 7.8: Finding a guide wire in the image of Fig. 7.7. Responses to high-noise Volterra filters (Result 1) for a direction process-based CIRF ($\epsilon = 10.5, \zeta = 0, \lambda = 100$, in all cases) with $\sigma_\kappa = 1/10$ and 64 directions (left), and with $\sigma_\kappa = 1/5$ and 32 directions (right). Observe how the cubic filter at $\sigma_\kappa = 1/10$ enhances the guide wire. To produce these two dimensional images from the actual discrete (x, y, θ) -space responses, the maximum over θ was taken.

taken. Observe how the many bogus responses in the thresholded local responses are not present in the Volterra results.

To test the biased CIRF posterior mean approximation filter, we applied it to an example we saw earlier in Fig. 7.5. We solved the coupled nonlinear system (5.32) and (5.33) for Gaussian noise and with the blur operator B simply taking an integral over θ . In the continuation method for solving the system, we started with $\epsilon = 0.58$ and let it grow until $\epsilon = 1.6$. We chose homogeneous initial and final weightings μ and ν , respectively, and set $\bar{\mathcal{N}} = 1$. The results (Fig. 7.10) show how effective the nonlinearity in the coupled system is at suppressing the weak responses away from the true contours.

7.2 Contour Filtering in $\mathbb{R}^2 \times \mathbb{S} \times \mathbb{R}$ with Curvature

In addition to MMSE CIRF Volterra filters for the direction process, we performed the analogous filtering for the curvature process in (discrete) $\mathbb{R}^2 \times \mathbb{S} \times \mathbb{R}$. For our first example, we considered a blood cell image (Fig. 7.11, top). To illustrate robustness, noise was added to a small portion of the image that contained two cells (top left), and was processed with the logical/linear edge operator at the default settings. Inputs for the curvature process were produced by copying over the κ -dimension, as curvature is not directly measured in the image. The result was first filtered using the CIRF posterior mean based on the direction process (top center). Despite using two very different bounds on curvature, the direction-based filtering cannot close the blood cell boundaries appropriately. In contrast, the CIRF posterior mean with the curvature process (top right) was more effective at forming a complete boundary. To illustrate in more detail, we plotted the filter responses for the direction-based filter at $\sigma_\kappa = 0.025$ for 8 of the 32 discrete directions in the middle of Fig. 7.11. The brightness in

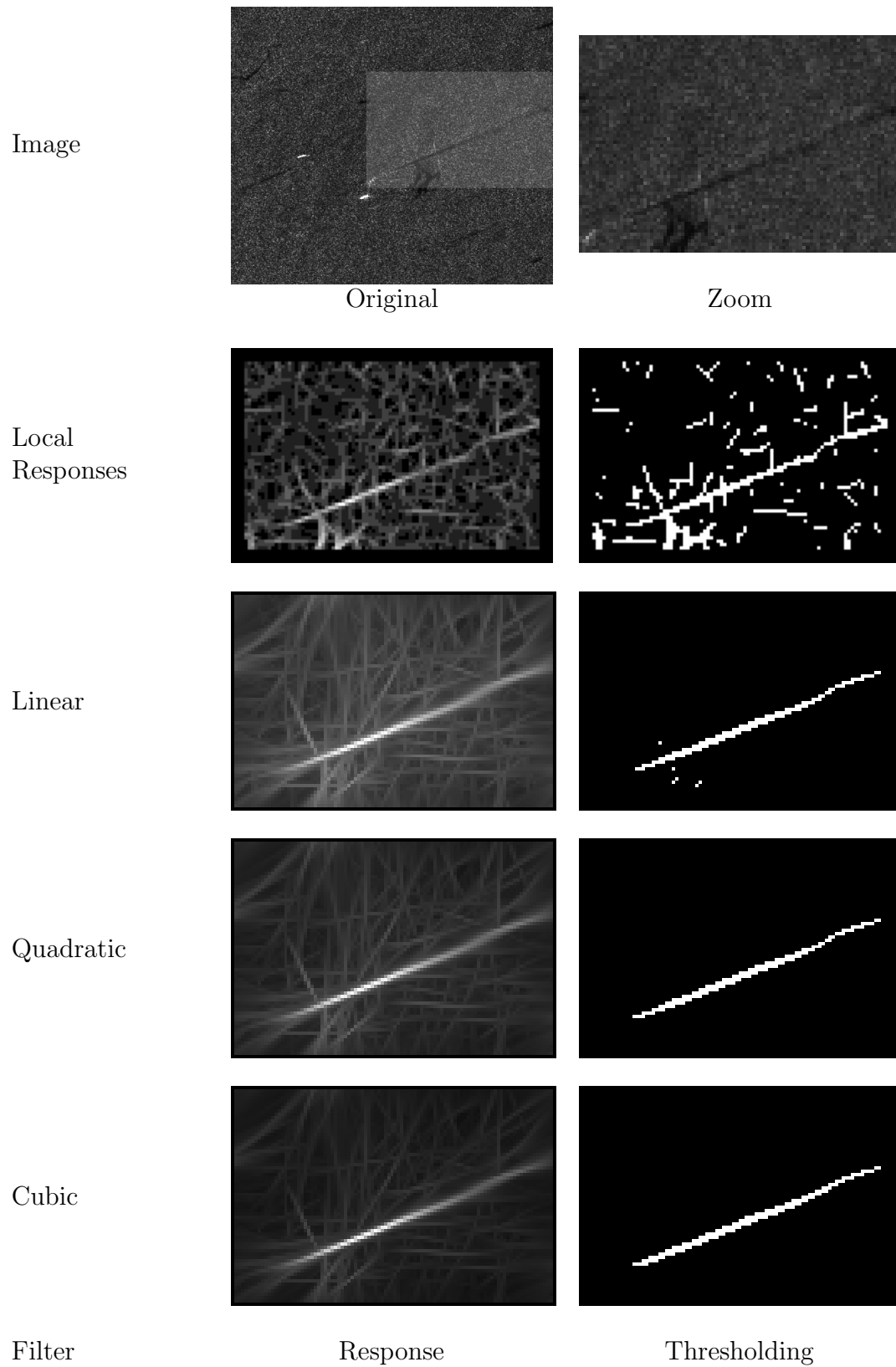


Figure 7.9: Finding a ship's wake (see text for explanation).

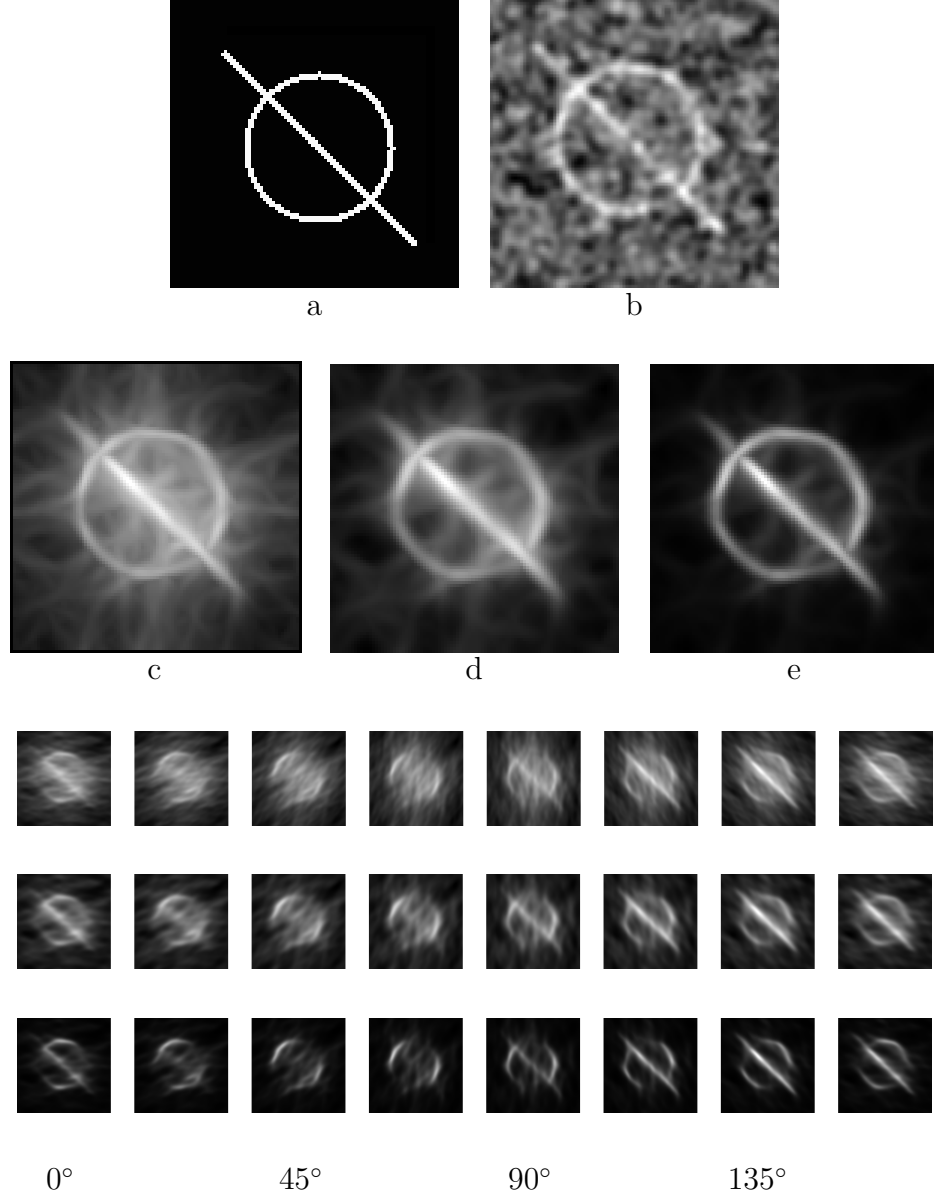


Figure 7.10: Filtering with the biased CIRF posterior mean approximation. (a) Original image, (b) noisy version copied over θ to form input m to filter, (c) cubic Volterra result ($\epsilon = 10.5, \zeta = 0, \sigma_\kappa = 1/3$, with 16 directions), (d) biased CIRF result at start of continuation method ($\epsilon = 0.58$), (e) final biased CIRF result at end of continuation method ($\epsilon = 1.6$). Bottom three rows are (from top to bottom): response over θ for cubic Volterra filter; for biased CIRF approximation at start of continuation; and for biased CIRF approximation at the end of the continuation method. Observe the striking degree of orientation selectivity in the biased CIRF approximation end result.

each of the 8 sub-images is proportional to the response for that particular direction as a function of position (x, y) . Observe the over-straightening effect shown by the elongated responses. The curvature filter responses were plotted as a function of direction and curvature (bottom). Despite the input having been constant as a function of curvature, the result shows *curvature selectivity*. Indeed, one can clearly see in the $\kappa > 0$ row (Fig. 7.11, bottom) that the boundary of the top left blood cell is traced out in a counter-clockwise manner. In the $\kappa < 0$ row, the same cell is traced out in the opposite manner. (Since the parameterization of the curve is lost when forming its image, we cannot know which way the contour was traversed; our result is consistent with both ways.) The response for the lower right blood cell was somewhat weaker but qualitatively similar. Unlike the direction-only process, the curvature process can effectively deal with highly curved contours.

For our next example, we took two sub-images of a low-contrast angiogram (top of Fig. 7.12; sub-images from left and top right of original). The first sub-image (top left) contained a straight structure, which was enhanced by our curvature-based CIRF filter (summed responses at top right). The distinct responses at separate directions and curvatures show curvature selectivity as well, since the straight curvature at 45° had the greatest response (center). The second sub-image (bottom left) of a loop structure also produced a reasonable filter response (bottom right); the individual responses (bottom) also show moderate curvature selectivity.

As argued in §2.4, the bug with a direction-only search cone would mis-track on a contour as curvature builds up. To make this point computationally, we considered an image of an Euler spiral extending from a straight line segment (Fig. 7.13, top left).² Observe that the contour curvature begins at zero (straight segment)

²We used formula (16.7) of Kimia et al [58], and created the plot in Mathematica with all parameters 0, except $\gamma = 0.1$ (Kimia et al's notation). The resulting plot was grabbed, combined with a line segment, blurred with a Gaussian, and then subsampled.

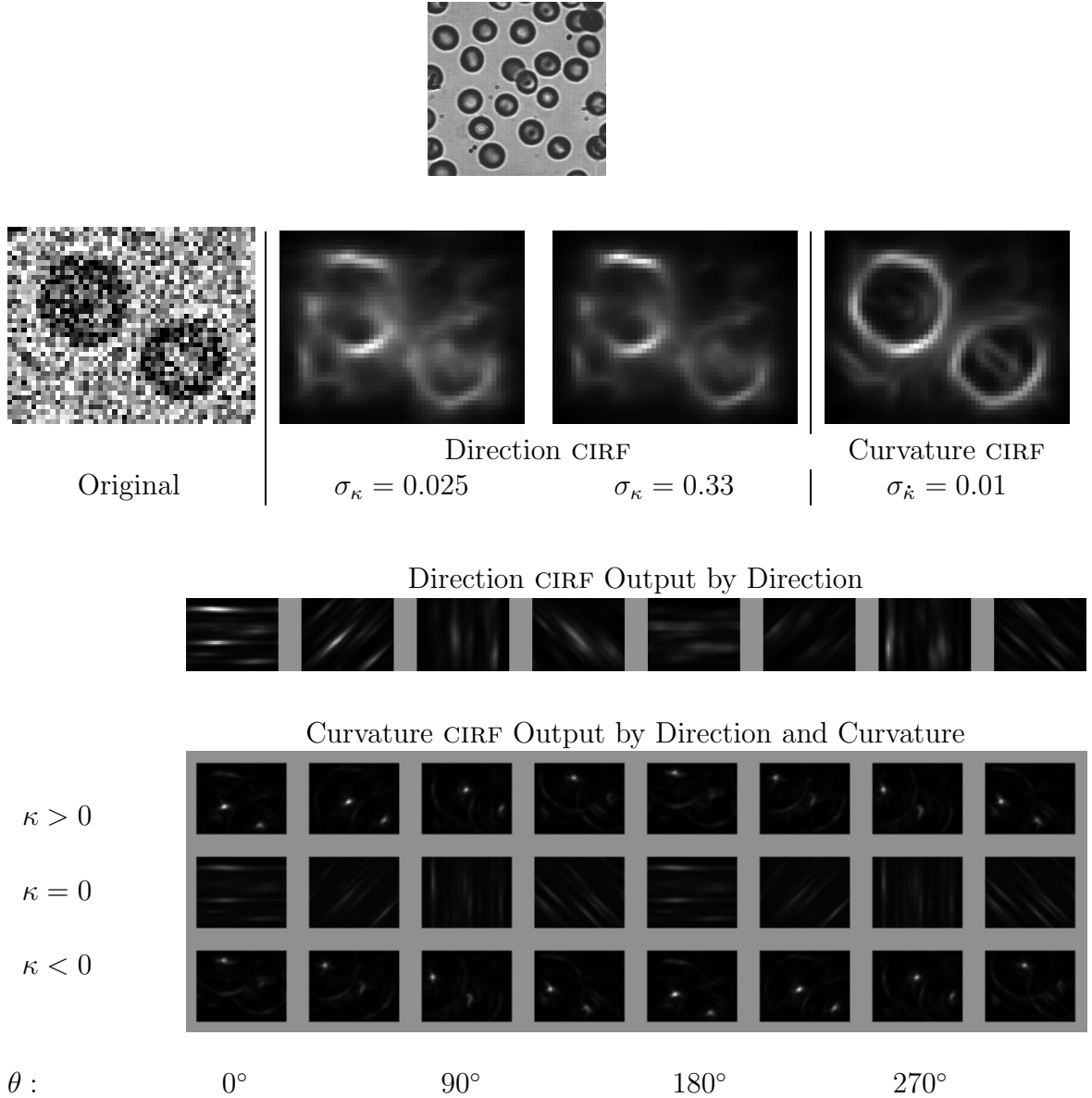


Figure 7.11: Curvature filtering of a blood cell image (see text).

and then builds up gradually. To produce a 3-dimensional input to our direction-based filter, this original (2-dimensional) image \mathbf{I} was copied to all directions (i.e., $m(x, y, \theta) = \mathbf{I}(x, y)$, for all θ). Similarly, the image was copied to all directions and curvatures to produce a 4-dimensional input to the curvature-based filter. The 2-dimensional outputs were produced by taking, at each position (x, y) , the maximum response over all directions θ (for the direction-based filtering) or over all directions

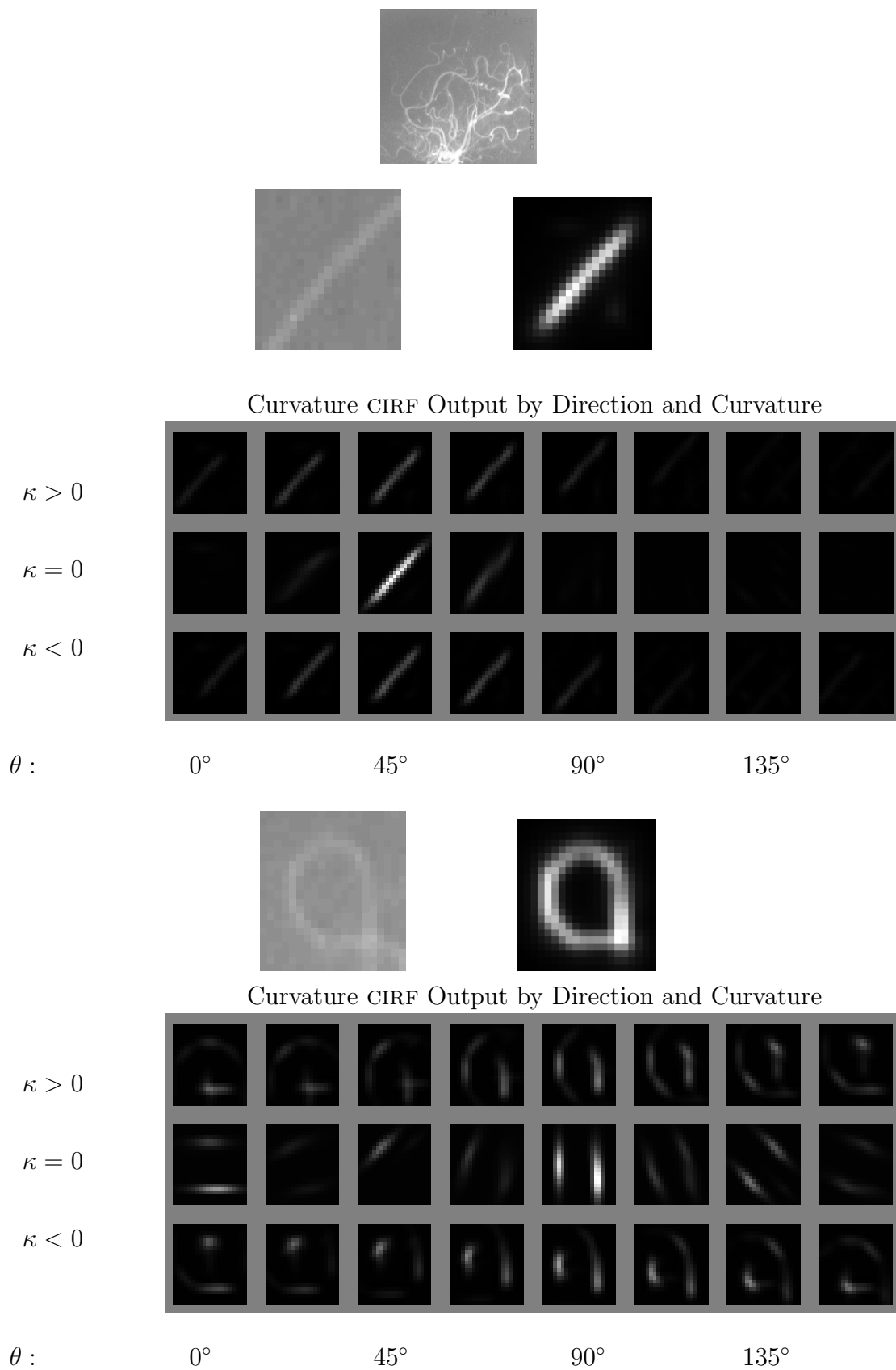


Figure 7.12: Curvature filtering for an angiogram (see text).

θ and curvatures κ (for the curvature-based filtering). The direction-based CIRF posterior mean (with parameters $\sigma_\kappa = 0.025$, $\lambda = 10$, and 64 directions) was computed (left, rows (a) and (b)), showing an undesirable reduction in response as curvature built up. The curvature-based CIRF posterior mean (left, rows (c) and (d), with parameters $\sigma_\kappa = 0.05$, $\lambda = 10$, 64 directions, and 7 curvatures $(0, \pm 0.05, \pm 0.1, \pm 0.15)$) shows strong response even at the higher curvature portions of the contour. To test robustness, 0-mean Gaussian noise of standard deviation 0.5 was added (top right) to the image (0 to 1 was the signal range before adding noise). The results (right, rows (a)-(d)) show that the curvature-based filter performs better in high curvature regions despite noise.

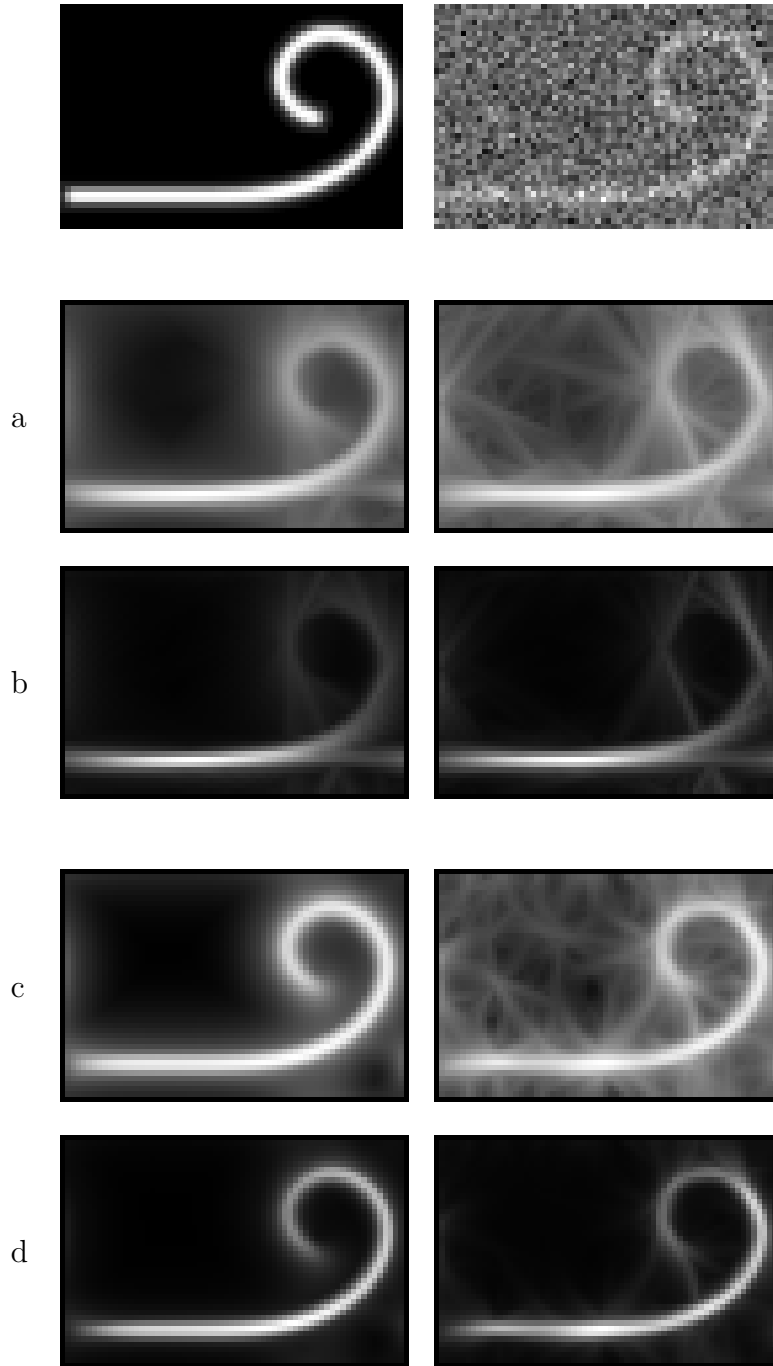


Figure 7.13: Euler spiral images without noise (top left) and with noise (top right) are processed with high-noise MMSE Volterra CIRF filters based on the direction process (linear and cubic filter responses in rows (a) and (b), respectively) and the curvature process (linear and cubic filter responses in row (c) and (d), respectively). Notice that the direction process CIRF result tends to repress the contour signal at high curvatures, while the curvature process has more consistent performance, even at higher curvatures. See text for details.

Chapter 8

Conclusion

In this thesis we have introduced a field structure intermediate between contours and the images that suggest them. This curve indicator random field was studied for Markov process contour models, leading to a practical description of its joint statistical structure via a moment generating functional. By relating contours to images, the model provides a way to verify whether it is in agreement with natural image structure; in this way a role for curvature was found, motivating our introduction of a Markov process in contour curvature. We believe that the CIRF is so basic that future work on visual contours will have difficulty avoiding it.

Enhancing contour structure in noisy images was characterized as MMSE estimation of an underlying CIRF, and a variety of nonlinear filters were provided for computing this posterior mean estimate. The Volterra filters emphasized high-noise situations, and results on synthetic and real images verified this. The more powerful reaction-diffusion-convection equation was provided for dealing with blur more accurately. We note that this PDE was not derived as an (explicit) Euler-Lagrange equation; instead, it arose indirectly via the moment generating functional of the biased CIRF.

Besides enabling explicit calculations, the CIRF provides a handle on a rather unpleasant aspect of contours: their global structure. In particular, the square of the CIRF measures contour intersections. We note that this was not suggested by studying the CIRF prior itself; instead it arose as a quadratic term in the likelihood. In hindsight, however, it can just as easily be incorporated into the CIRF prior to discourage intersections (see §5.2.3). No current scheme for contour inference has a place for these interactions, although it is not surprising that dealing with multiple contours must eventually confront the intersection problem.

8.1 Future Directions

The work in this thesis can be both focused and broadened in a number of ways. On the practical side, the estimation of the parameters used in the CIRF filters would be convenient for further automating their use. More interesting is the issue of learning the Markov process contour model. At its simplest, this could be viewed as a model selection problem, perhaps for choosing between the direction and curvature processes. More generally, one might like to learn the Markov process generator itself, using empirical edge correlations of training curve images, for example.

Even more important for applications may be the estimation of good observation models. We have outlined how the Geman and Jedynak scheme [35] of on-edge and off-edge distributions can be incorporated. Clearly more realism is desirable; of highest priority is probably blur scale; currently, our scheme is implemented at one scale with the direction and curvature processes.

Another immediate issue is occlusion, which is a fundamental example of contour intersection. Can this be formalized with a quadratic log-likelihood observation model? Possibly, because occlusion suggests a competitive interaction, where the

presence of a contour would discourage other contours on one side (the occluding side). Another aspect of interaction is texture flow [49], where approximately parallel nearby contours merge into a flow field of local orientations.

Interaction is also relevant for branching structures, such as vascular structures in the human body. One can formalize this using branching Markov process (superprocesses). This would be particularly important for medical imaging.

Both closure and symmetry can be thought of as interactions of another form, by representing the contour with its skeleton or symmetry set. Such a model would allow us to characterize regional properties of the shape, such as symmetry of two portions of boundary. Even though symmetry is not a local property of the boundary, it is a local property of the skeleton. We can construct a branching Markov process for such a medial axis description that can specify symmetry information through the radius function and axis location (and derivatives). Two local constraints that must be maintained in the process are: (1) the radius of curvature of the skeleton is bounded below by the radius and (2) the rate of change of radius is less than unity (local growth). Closure (of the boundary induced by the skeleton) is maintained by construction. The key complication introduced here is the need for a CIRF representation for the boundary and another for the skeleton. For consistency, these two CIRFs will have to be coupled so that they correspond to the same shapes.

There are a number of interesting technical issues related to the CIRF. For example, what is the role of the (approximate) self-avoidingness of the Markov process models, and how does this relate to the intersection terms in the likelihood? While functions of Markov processes are generally not themselves Markovian, is there even an approximate sense in which the CIRF, as a random field, is Markovian? How is this related to contour self-avoidingness? On a simpler note, it would be useful to

know how to solve for the initial and final weightings on contour endpoints in terms of given endpoint marginal distributions (Remark 1).

Mathematically, it would also be useful to know some basic properties of the filtering equations we have provided. Do solutions exist? Are they unique? For the Volterra filters these questions are trivial, but for the biased CIRF approximation these questions are more subtle.

In a more speculative direction, one may also extend the framework presented here beyond contours, to submanifolds of dimension greater than one. Inspired by applications of relaxation labeling [105], surfaces would be the most important immediate extension; one could define an analogous *surface indicator random field* (SIRF) to support medical imaging applications, for example. One goal would be to estimate the SIRF that corresponds to the intestinal wall or the surface of a brain from CAT or MRI images.

It is the hope of this work that it encourages the field of computer vision to consider taking low-level information, aided by generic geometrical priors, more seriously as a foundation for general vision systems.

Bibliography

- [1] P. Alper. A consideration of discrete volterra series. *IEEE Transactions on Automatic Control*, 10:322–327, 1965.
- [2] H. C. Andrews and B. R. Hunt. *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs, NJ, 1977.
- [3] J. August and S. W. Zucker. The curve indicator random field: curve organization via edge correlation. In K. Boyer and S. Sarkar, editors, *Perceptual Organization for Artificial Vision Systems*, pages 265–288. Kluwer Academic, Boston, 2000.
- [4] J. August and S. W. Zucker. The moments of the curve indicator random field. In *Proceedings of the 2000 Conference on Information Sciences and Systems*, volume 1, pages WP5–19–WP5–24, Princeton, NJ, March 2000.
- [5] J. August and S. W. Zucker. Organizing curve elements with an indicator random field on the (unit) tangent bundle. In *IEEE Workshop on Perceptual Organization in Computer Vision*, Corfu, September, 1999.
- [6] P. Belhumeur. A bayesian-approach to binocular stereopsis. *IJCV*, 19(3):237–260, August 1996.

- [7] J. Biemond, R. L. Lagendijk, and R. M. Mersereau. Iterative methods for image deblurring. *Proc. IEEE*, 78(5):856–883, 1990.
- [8] J. Biemond, J. Rieske, and J. J. Gerbrands. A fast kalman filter for images degraded by both blur and noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31:1248–1256, 1983.
- [9] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.
- [10] P. Bondon, P. L. Combettes, and B. Picinbono. Volterra filtering and higher order whiteness. *IEEE Transactions on Signal Processing*, 43(9):2209–2212, 1995.
- [11] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.
- [12] S. Casadei and S. Mitter. Hierarchical image segmentation – Part I: Detection of regular curves in a regular graph. *International Journal of Computer Vision*, 27(3):71–100, 1998.
- [13] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, Belmont, CA, 1990.
- [14] D. E. Catlin. *Estimation, Control, and the Discrete Kalman Filter*. Springer, New York, 1989.
- [15] I. J. Cox, J. M. Rehg, and S. Hingorani. A Bayesian Multiple-Hypothesis Approach to Edge Grouping and Contour Segmentation. *International Journal of Computer Vision*, 11:5–24, 1993.

- [16] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160–1169, 1985.
- [17] A. Desolneux, L. Moisan, and J.-M. Morel. Meaningful alignments. *International Journal of Computer Vision*, 40(1):7–23, 2000.
- [18] A. Desolneux, L. Moisan, and J.-M. Morel. Edge detection by helmholtz principle. *Journal of Mathematical Imaging and Vision*, 14:271–284, 2001.
- [19] A. Dobbins, S. W. Zucker, and M. S. Cynader. Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329(6138):438–441, 1987.
- [20] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.
- [21] J. S. Duncan and T. Birkholzer. Reinforcement of linear structure using parametrized relaxation labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(5):502–515, 1992.
- [22] E. B. Dynkin. Markov processes as a tool in field theory. *Journal of Functional Analysis*, 50:167–187, 1983.
- [23] E. B. Dynkin. Gaussian and non-gaussian fields associated with markov processes. *Journal of Functional Analysis*, 55:344–376, 1984.
- [24] E. B. Dynkin. Local times and quantum fields. In *Seminar on Stochastic Processes, 1983*, pages 69–83, Boston, 1984. Birkh auser.

- [25] J. Elder and A. Krupnik. Contour grouping with strong prior models. In *The 3rd Workshop on Perceptual Organization in Computer Vision*, Vancouver, 2001.
- [26] J. H. Elder. *The Visual Computation of Bounding Contours*. PhD thesis, McGill University, 1995.
- [27] R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models: Estimation and Control*. Springer, 1994.
- [28] L. Euler. *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes*. Lausanne, 1744.
- [29] D. J. Field, A. Hayes, and R. Hess. Contour integration by the human visual system: Evidence for an a local ‘association field’. *Vision Research*, 33:173–193, 1993.
- [30] P. J. Fitzsimmons and J. Pitman. Kac’s moment formula and the feynman-kac formula for additive functionals of a markov process. *Stochastic Processes and their Applications*, 79:117–134, 1999.
- [31] L. Florack. *The Syntactical Structure of Scalar Images*. PhD thesis, Utrecht University, 1993.
- [32] L. Florack, B. ter Haar Romeny, M. Viergever, and J. Koenderink. The gaussian scale-space paradigm and the multiscale local jet. *International Journal of Computer Vision*, 18:61–75, 1996.
- [33] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever. Scale and the differential structure of images. *Image and Vision Computing*, 10(6):376–388, August 1992.

- [34] W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724, 2001.
- [35] D. Geman and B. Jedynek. An active testing model for tracking roads in satellite images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1):1–14, 1996.
- [36] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.
- [37] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer Academic, Dordrecht, 1995.
- [38] U. Grenander and M. I. Miller. Representations of knowledge in complex-systems. *Journal of the Royal Statistical Society Series B-Methodological*, 56(4):549–603, 1994.
- [39] G. Guy and G. Medioni. Inferring global perceptual contours from local features. *International Journal of Computer Vision*, 20(1/2):113–133, 1996.
- [40] E. R. Hancock and J. Kittler. Edge-labeling using dictionary-based relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):165–181, 1990.
- [41] C. W. Helstrom. *Probability and Stochastic Processes for Engineers*. Macmillan, New York, 1991.
- [42] L. Herault and R. Horaud. Figure-ground discrimination: A combinatorial optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):899–914, 1993.

- [43] B. K. P. Horn. The Curve of Least Energy. *ACM Transactions on Mathematical Software*, 9:441–460, 1983.
- [44] J. Huang and D. Mumford. Statistics of natural images and models. In *Proceedings, Computer Vision and Pattern Recognition*, pages 541–547, 1999.
- [45] D. H. Hubel and T. N. Wiesel. Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond. B*, 198:1–59, 1977.
- [46] M. H. Hueckel. An operator which locates edges in digitized pictures. *Journal of the Association for Computing Machinery*, 18(1):113–125, 1971.
- [47] R. Hummel and S. W. Zucker. On the foundations of relaxation labeling processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:267–287, 1983.
- [48] M. Isard and A. Blake. Condensation: conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.
- [49] L. A. Iverson. *Toward Discrete Geometric Models for Early Vision*. PhD thesis, McGill University, Montreal, 1994.
- [50] L. A. Iverson and S. W. Zucker. Logical/linear operators for image curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):982–996, 1995.
- [51] S. N. Kalitzin, B. M. ter Haar Romeny, and M. A. Viergever. Invertible orientation bundles on 2d scalar images. In *Proc. Scale-Space '97*, LICS, pages 77–88. Springer, 1997.
- [52] G. Kanizsa. *Organization in Vision*. Praeger, New York, 1979.

- [53] H. J. Kappen and F. B. Rodriguez. Efficient learning in boltzmann machines using linear response theory. *Neural Computation*, 10:1137–1156, 1998.
- [54] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, pages 321–331, 1988.
- [55] T. Kato. *A Short Introduction to Perturbation Theory for Linear Operators*. Springer, New York, 1982.
- [56] C. T. Kelly. *Iterative Methods for Linear and Nonlinear Equations*. SIAM, 1995.
- [57] R. Z. Khas'minskii. On positive solutions of the equation $uu+vu$. *Theory of Probability and Its Applications*, 4(3):309–318, 1959.
- [58] B. B. Kimia, I. Frankel, and A.-M. Popescu. Euler spiral for shape completion. In K. Boyer and S. Sarkar, editors, *Perceptual Organization for Artificial Vision Systems*, pages 289–309. Kluwer Academic, Boston, 2000.
- [59] J. J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.
- [60] J. J. Koenderink and W. Richards. Two-dimensional curvature operators. *J. Opt. Soc. Am. A*, 5(7):1136–1141, 1988.
- [61] K. Koffka. *Principles of Gestalt Psychology*. Harcourt, Brace & World, New York, 1963.
- [62] T. Koh and E. J. Powers. Second-order volterra filtering and its application to nonlinear system identification. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(6):1445–1455, 1985.

- [63] S. Konishi and A. L. Yuille. Statistical cues for domain specific image segmentation with performance analysis. In *Proceedings, Computer Vision and Pattern Recognition*, pages 125–132, 2000.
- [64] N. Kruger. Collinearity and parallelism are statistically significant second-order relations of complex cell responses. *Neural Processing Letters*, 8:117–129, 1998.
- [65] P. Kube. Polynomial shift-invariant operators for texture segmentation. In *CVPR88*, pages 100–104, 1988.
- [66] V. P. Leonov and A. N. Shiryaev. On a method of calculation of semi-invariants. *Theory of Probability and Its Applications*, 4(3):319–329, 1959.
- [67] S. Z. Li. *Markov Random Field Modelling in Computer Vision*. Springer, Tokyo, 1995.
- [68] S. Z. Li. Modeling image analysis problems using markov random fields. In C. R. Rao and D. N. Shanbhag, editors, *Stochastic Processes: Modeling and Simulation*, volume 20 of *Handbook of Statistics*. Elsevier Science, 2001.
- [69] W. Maass and E. D. Sontag. Neural systems as nonlinear filters. *Neural Computation*, 12(8):1743–1772, 2000.
- [70] S. Mahamud, K. K. Thornber, and L. R. Williams. Segmentation of salient closed contours from real images. In *Proceedings, International Conference on Computer Vision*, 1999.
- [71] J. L. Marroquin. *Probabilistic Solution of Inverse Problems*. PhD thesis, Massachusetts Institute of Technology, 1985.
- [72] J. L. Marroquin. A markovian random field of piecewise straight lines. *Biological Cybernetics*, 61:457–465, 1989.

- [73] A. Martelli. An application of heuristic search methods to edge and contour detection. *Comm. of the ACM*, 19(2):73–83, 1976.
- [74] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the Eighth International Conference on Computer Vision*, volume II, pages 416–423, 2001.
- [75] P. McCullagh. Tensor notation and cumulants of polynomials. *Biometrika*, 71(3):461–476, 1984.
- [76] P. McCullagh. *Tensor Methods in Statistics*. Monographs in Statistics and Applied Probability. Chapman and Hall, Cambridge, England, 1987.
- [77] D. Middleton. *An Introduction to Statistical Communication Theory*. Peninsula Publishing, Los Altos, CA, 1987.
- [78] D. A. Miller and S. W. Zucker. Copositive-plus lemke algorithm solves polymatrix games. *Operations Res. Lett.*, 10:285–290, 1991.
- [79] R. Molina, A. del Olmo, J. Perea, and B. D. Ripley. Bayesian deconvolution in optical astronomy. *The Astronomical Journal*, 103(2):666–675, 1992.
- [80] U. Montanari. On the optimum detection of curves in noisy pictures. *Comm. ACM*, 14:335–345, 1971.
- [81] M. C. Morrone and D. C. Burr. Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society of London B*, 235:221–245, 1988.
- [82] D. Mumford. *Algebraic Geometry and Its Applications*, chapter Elastica and Computer Vision, pages 491–506. Springer-Verlag, 1994.

- [83] D. Mumford and B. Gidas. Stochastic models for generic images. *Quart. Appl. Math.*, 59(1):85–111, 2001.
- [84] C. L. Nikias and A. P. Petropulu. *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*. Prentice Hall, Englewood Cliffs, NJ, 1993.
- [85] M. Nitzberg, D. Mumford, and T. Shiota. *Filtering, Segmentation and Depth*. Springer-Verlag, Berlin, 1993.
- [86] J. R. Norris. *Markov Chains*. Cambridge UP, 1997.
- [87] R. Nowak and B. V. Veen. Tensor product basis approximations for volterra filters. *IEEE Transactions on Signal Processing*, 44(1):36–50, 1996.
- [88] K. Okajima. A model visual cortex incorporating intrinsic horizontal neuronal connections. *Neural Networks*, 9(2):211–222, 1996.
- [89] P. Parent and S. W. Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):823–839, August 1989.
- [90] G. Parisi. *Statistical Field Theory*. Perseus Books, Reading, MA, 1998.
- [91] C. Peterson and J. R. Anderson. A mean field theory learning algorithm for neural networks. *Complex Systems*, 1:995–1019, 1987.
- [92] B. Picinbono and Y. Gu. Mean square estimation and projections. *Signal Processing*, 19:1–8, 1990.
- [93] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, September 1985.

- [94] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer, New York, 1988.
- [95] W. K. Pratt. *Digital Image Processing*. Wiley, 1978.
- [96] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.
- [97] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–285, 1989.
- [98] L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, pages 4–15, January 1986.
- [99] J. I. Richards and H. K. Youn. *Theory of Distributions: a non-technical introduction*. Cambridge U P, Cambridge, UK, 1990.
- [100] L. Roberts. Machine perception of 3-dimensional solids. In J. Tippett, editor, *Optical and Electro-Optical Information Processing*. MIT Press, Cambridge, MA, 1965.
- [101] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes, and Martingales*, volume 1. Wiley, Chirchester, 2nd edition, 1994.
- [102] A. Rosenfeld, R. A. Hummel, and S. W. Zucker. Scene labeling by relaxation operations. *IEEE Transactions on Systems, Man, and Cybernetics*, 6(6):420–433, 1976.
- [103] D. L. Ruderman. *Natural Ensembles and Sensory Signal Processing*. PhD thesis, U. C. Berkeley, 1993.

- [104] D. L. Ruderman. The statistics of natural images. *Network: Computation in Neural Systems*, 5:517–548, 1994.
- [105] P. T. Sander and S. W. Zucker. Inferring surface trace and differential structure from 3-d images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):833–853, 1990.
- [106] L. K. Saul, T. Jaakkola, and M. Jordan. Mean field theory for sigmoid belief networks. *Journal of Artificial Intelligence Research*, 4:61–76, 1996.
- [107] M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley, New York, 1980.
- [108] A. Sha’ashua and S. Ullman. Structural saliency: The detection of globally salient structures using a locally connected network. In *Proceedings, Second International Conference on Computer Vision*, 1988.
- [109] E. Sharon, A. Brandt, and R. Basri. Completion energies and scale. In *Proc. CVPR*, pages 884–890, 1997.
- [110] M. Sigman, G. A. Cecchi, C. D. Gilbert, and M. O. Magnasco. On a common circle: Natural scenes and gestalt rules. *Proceedings of the National Academy of the Sciences*, 98(4):1935–1940, 2001.
- [111] E. P. Simoncelli. Modeling the joint statistics of images in the wavelet domain. In *Proc. SPIE 44th Annual Meeting*, volume 3813, 1999.
- [112] T. P. Speed. Cumulants and partition lattices. *Australian Journal of Statistics*, 25(2):378–388, 1983.
- [113] P. Swerling. A proposed stagewise differential correction procedure for satellite tracking and predictions. *J. Astronaut. Sci.*, 6:46–59, 1959.

- [114] T. Tanaka. Mean-field theory of boltzmann machine learning. *Physical Review E*, 58(2):2302–2310, 1998.
- [115] T. Tanaka. Information geometry of mean-field approximation. *Neural Computation*, pages 1951–1968, 2000.
- [116] K. Thornber and L. Williams. Analytic solution of stochastic completion fields. *Biological Cybernetics*, 75:141–151, 1996.
- [117] K. K. Thornber and L. R. Williams. Orientation, scale, and discontinuity as emergent properties of illusory contour shape. In *Neural Information Processing Systems (NIPS '98)*, Denver, CO, 1998.
- [118] K. K. Thornber and L. R. Williams. Characterizing the distribution of completion shapes with corners using a mixture of random processes. *Pattern Recognition*, 33:543–553, 2000.
- [119] S. Thrun, D. Fox, W. Burgard, and F. Dellaert. Robust monte carlo localization for mobile robots. *Artificial Intelligence*, 2001.
- [120] A. Turiel, G. Mato, N. Parga, and J.-P. Nadal. Self-similarity properties of natural images resemble those of turbulent flows. *Physical Review Letters*, 80(5):1098–1101, 1998.
- [121] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Physical Review*, 36:823–841, 1930.
- [122] S. Ullman. Filling-in gaps: The shape of subjective contours and a model for their generation. *Biological Cybernetics*, 25:1–6, 1976.
- [123] A. van der Schaaf. *Natural Image Statistics and Visual Processing*. PhD thesis, Groningen University, The Netherlands, 1998.

- [124] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. Wiley, New York, 1960.
- [125] L. Williams and D. Jacobs. Local parallel computation of stochastic completion fields. *Neural Computation*, 9(4):859–881, 1997.
- [126] L. Williams and D. Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation*, 9(4):837–858, 1997.
- [127] L. Williams, T. Wang, and K. Thornber. Computing stochastic completion fields in linear-time using a resolution pyramid. In *Proc. of 7th Intl. Conf. on Computer Analysis of Images and Patterns*, Kiel, Germany, 1997.
- [128] G. Winkler. *Image Analysis, Random Fields, and Dynamic Monte Carlo Methods*. Springer, 1995.
- [129] A. Witkin, D. Terzopoulos, and M. Kass. Signal matching through scale space. *International Journal of Computer Vision*, 1(2):133–144, 1987.
- [130] J. W. Woods. Two-dimensional discrete markovian fields. *IEEE Transactions on Information Theory*, 18(2):232–240, 1972.
- [131] J. W. Woods and V. K. Ingle. Kalman filtering in two dimensions: Further results. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(2):188–197, 1981.
- [132] J. W. Woods and C. H. Radewan. Kalman filtering in two dimensions. *IEEE Transactions on Information Theory*, 23(4):473–482, 1977.
- [133] A. L. Yuille and J. M. Coughlan. Fundamental bounds of bayesian inference: Order parameters and phase transitions for road tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2):160–173, 2000.

- [134] C. Zetsche, E. Barth, and B. Wegmann. The importance of intrinsically two-dimensional image features in biological vision and picture coding. In A. B. Watson, editor, *Digital Images and Human Vision*, pages 109–138. MIT, Cambridge, MA, 1993.
- [135] S.-C. Zhu. Embedding gestalt laws in markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1170–1187, 1999.
- [136] S. C. Zhu and D. Mumford. Grade: Gibbs reaction and diffusion equations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11), 1997.
- [137] S. C. Zhu, Y. N. Wu, and D. Mumford. Minimax entropy principle and its application to texture modeling. *Neural Computation*, 9(8):1627–1660, 1997.
- [138] S. W. Zucker, A. Dobbins, and L. Iverson. Two stages of curve detection suggest two styles of visual computation. *Neural Computation*, 1:68–89, 1989.
- [139] S. W. Zucker, R. Hummel, and A. Rosenfeld. An application of relaxation labelling to line and curve enhancement. *IEEE Trans. Computers*, C-26:393–403, 922–929, 1977.
- [140] J. Zweck and L. R. Williams. Euclidean group invariant computation of stochastic completion fields using shiftable-twistable functions. In *Proc. of the 6th European Conference on Computer Vision*, Dublin, Ireland, 2000.

Index

- CIRF, 31
- MAP, 65
- MGF, 39
- MMSE, 67
- PDE, 22
- approximately self-avoiding Markov processes, 100, 102
- co-circularity, 54
- collinearity, 54
- correlation, 53
- correlation length, 54
- covariance, 51
- cumulant tensor representation, 85
- curvature, 10, 54
- curve enhancement, 77
- curve indicator random field, 31
- density, 91
- direction, 9
- edge correlations, 53
- edge de-blurring, 78
- elastica, 19
- error rate, 66
- generalized cumulants, 85
- generator, 16
- geometry, 54
- good continuation, 1
- grouping, 1
- homogeneous random field, 53
- ideal edge/line map, 4
- ideal edge/line map desiderata, 4
- image, 1
- image of contours, 3
- image of curves, 3
- Kronecker delta, 97
- lift, lifting, 11
- line process, 4
- Markov process, 10
- maximum a posteriori, 65
- orientation, 9

oriented Wiener filter, 78

partial differential equation, 22

perceptual organization, 1

performance evaluation, 138

posterior mode, 65

relative residual, 127

relaxation labeling, 54

saliency, 4

semantic gap, 1

signal-to-symbol transition, 1

sink field, 114

site-level error rate, 66

source field, 114

summation convention, 84

tangent bundle, 78

tensor, 84

voting kernel, 77