CLUTTER 17: NON-DIFFEOMORPHIC BACKGROUND DEFORMATIONS[1]

Ulf Grenander, Brown University

Division of Applied Mathematics

Providence, 02912 RI

May, 2002

1 Some Background Deformations in Pattern Theory

To create knowledge representations for inference in Computer Vision the following set up has been used successfully, see for example Grenander(1993), Grenander-Miller(1999) where many references can be found to related work. Start from a collection of templates $I_1^{template}, I_2^{template}, ...I_\alpha^{template}, ...$ and a similarity group S acting upon the background space X on which the templates are defined. On S is defined a prior probability measure P that represents the variability of the images that we observe. Combining templates via a connector graph σ and identifying the resulting image by an identification rule R we get an image

$$I = R[\sigma(s_1 I_{i_1}, s_2 I_{i_2}, s_3 I_{i_3}, ...); s_i \in S \tag{1}$$

Here the background space X will be a rectangle in the plane, and R could be, for example, identifying sums or unions of images. It is convenient to talk about

[1]Supported by NSF DMS 00774276

X as a continuum, but when it comes to computing, to treat it as a discrete lattice. Therefore we shall use concepts like "derivative", "continuous", "curve", "boundary" both in their continuous and discrete versions.

The similarity group S has usually been assumed to be either a finite dimensional Lie group or the group of diffeomorphism over X, the latter meaning at least diffeomorphic with large probability w.r.t. P. An important case is when S is the additive group $\mathbf{R}^2$ and the probabilities are produced by a stochastic P.D.E.

$$Ls = e; s = (s_1, s_2) \in \mathbf{R}^2 \tag{2}$$

where L is a linear differential (or difference) operator, space invariant (constant coeeficients) of order $d = 2p$ with some boundary conditions BC that make it self-adjoint. We shall write

$$L = \sum_{\alpha,\beta=0}^{d} l_{\alpha,\beta} \frac{\partial^{\alpha+\beta}}{\partial^{\alpha} s_1 \partial^{\beta} s_2}; l = (l_{\alpha,\beta}) \tag{3}$$

Further we shalll let $e(\cdot)$ be the derivative of the 2D Wiener process in the plane. To make (1) meaningful L should be non-singular.

Of course the solution of (1) need not be diffeomoprphic a.s., but with appropriate choice of coeficients it can be made so with large probability. A sample solution with $d = 2$ can look like Figure 1

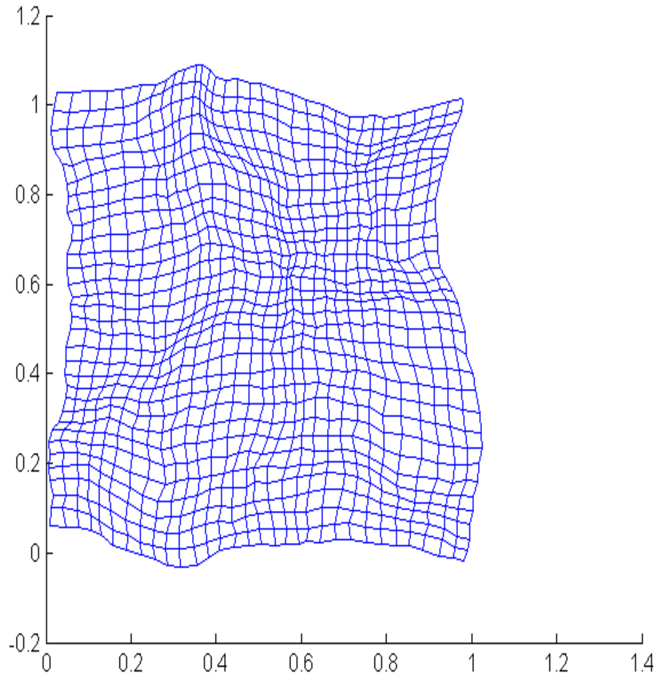

Figure 1

The inference can then be organized in terms of the likelihood function that takes the simple form

$$L(s) = \frac{constant}{\sqrt{J}} exp[-1/2Q(s)] \tag{4}$$

with the Jacobian

$$J = exp\{\sum_k log[\lambda_k]\} \tag{5}$$

where the λ's are the eigen values of the operator $Q = ll'$ and $Q(s)$ is the quadratic form in the vector (given as a matrix) l. The Toeplitz approximation gives

$$log[J] \asymp constant \times \int_{\mathbf{T}^2} log[R(\lambda)]d\lambda \tag{6}$$

integrated over the 2D torus, perhaps parametrized as $[-\pi, \pi)^2$, and with the Fourier transform

$$R(\lambda) = |\sum_{\alpha,\beta} i^{\alpha+\beta}\lambda_1^{\alpha}\lambda_2^{\beta}|^2 \tag{7}$$

For more details, see Lanterman,Grenander,Miller(2000).

2 Images Deformed by Non-Diffeomorphic Mappings

This is all very well, but if we encounter image ensembles with strong discontinuities, but still generated from continuous templates, then the above knowledge representation is inadequate.

To get some feeling for what sort of knowledge representation to choose for such partly discontinuous image ensembles, consider the picture in Figure 2

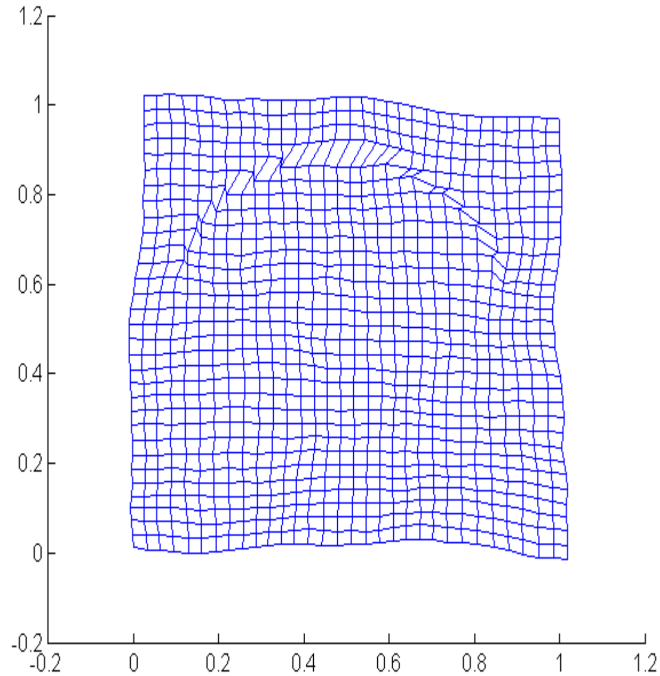

Figure 2

Here the discontinuities are produced along a circular arc, a cutset, and in Figure 3 we display *the horizontal and vertical line segments that have been cut by the disconuities*

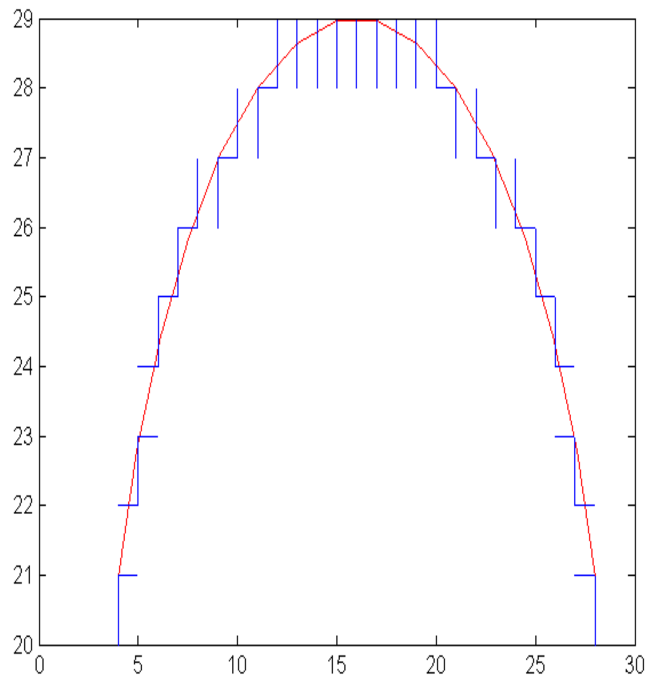

Figure 3

The likelihood, the conditional probability density for a fixed template, should describe how those discontinuities are likely to occur. The usual Markov random fields are not adequate; indeed the pertinent fact about the discontinuities is that they should occur at known locations.

Therefore it seems more promising, to build these expected curves of discontinuities into the likelihood. This will now be done and we shall use Gaussian measures to illustrate the idea of cutset processes. Let us remark, however, that the Gaussian assumption can be replaced by others.

REMARK. If the cutset forms a closed curve (in the discrete topology defined by neighborhood definition) it is clear that the random field inside the cutset is stochastically independent of the field outside so that no additional difficulty appears. In the opposite case, that seems to be the practical one, a new difficulty appears, which will be dealt with below.

Consider a partial stochastic difference equation

$$L(c)f = e \tag{8}$$

where the random fields f and e are defined on a $l \times l$ square lattice, e is i.i.d. $N(0, \sigma^2)$, say $\sigma = 1$, and $L(c)$ some non-singular linear operator that will be made to depend upon the cutset c, consisting of one or several smooth curves in $[0,1]^2$. Hence $f = L^{-1}(c)e$ has a Gaussian distribution with mean zero and the covariance operator $\sigma^2 L^{-2}(c)$.

We shall start with $\Delta + aI$ but modified by a term due to c, i.e. we define

$$L(c) = \Delta + aI - N(c) \tag{9}$$

Here Δ is the discrete Laplacian, corresponding to some suitable boundary conditions. The modifying term $N(c)$ shall be of the form

$$N(c) = \sum_{\overrightarrow{z,z'} cuts \quad c} \text{cut}(f; z, z') \tag{10}$$

where cut $(f; z, z')$ is an operator defined through the inner product

$$(\text{cut}(f : z, z'), g) = \frac{1}{4}[f(z)g(z) - f(z)g(z') - f(z')g(z) + f(z')g(z')]$$

for each segment $\overrightarrow{z, z'}$ that is cut by c. The segments can be either horizontal or vertical; see Figure 3.

The rationale behind this knowledge representation is that in images like the one in Figure 2 large discontinuities can occur at some specified locations but otherwise the field varies more slowly . Whether the latter occurs or not it is clear that (9) allows jumps around c, as we wish.

To synthesize such patterns is easy. We first simulate the i.i.d. e-field and then solve the Poisson equation (9) by classical relaxation (deterministic, not stochastic!). The value of a should be chosen very small. To discuss this we note, also for later use, that Δ has eigenvalues

$$\lambda_{st} = 1 - \frac{1}{2}\cos\frac{\pi s}{l} - \frac{1}{2}\cos\frac{\pi t}{l} \tag{11}$$

and the corresponding eigenfunctions

$$\psi_{st}(x, y) = \frac{2}{l}\sin\frac{\pi s x}{l}\sin\frac{\pi t y}{l}, \quad x, y = 1, \ldots, l \tag{12}$$

for $s, t = 1, \ldots, l$. Hence the smaller eigenvalues are, for large l,

$$\lambda_{st} \sim \frac{\pi^2}{4l^2}(s^2 + t^2)$$

so that $(\Delta + aI)^{-1}$ will have the larger eigenvalues approximately

$$(\lambda_{st} + a)^{-1} \sim \left[\frac{\pi^2}{4l^2}(s^2 + t^2) + a\right]^{-1}$$

If they vary slowly with s, t the field f will be too chaotic, but we want them to decrease fairly fast so that $a/(\frac{\pi^2}{4l^2} + a)$ should be small. Finally, to illustrate the way a cutset can appear we show a non-diffeomorphic mapping in Figure 4.

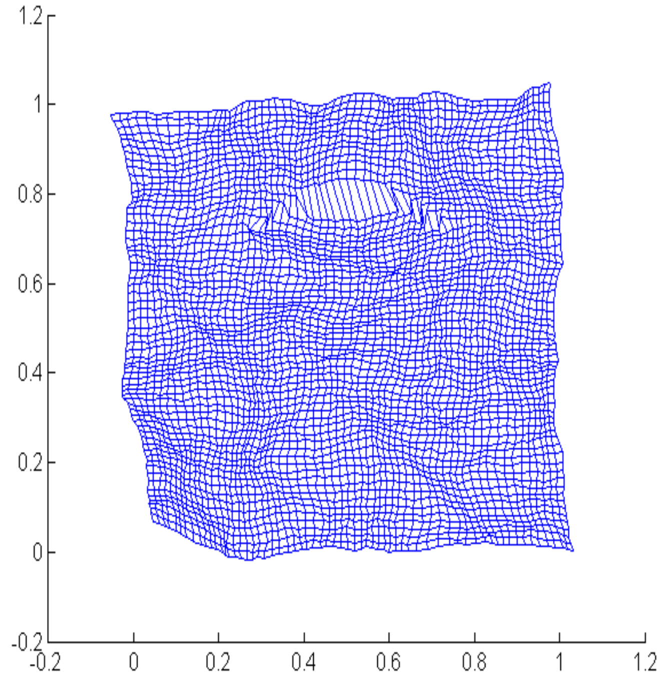

Figure 4

3 Perturbed Spectra of Covariance Operators for Cutset Processes

The likelihood function can be expressed as

$$\mathcal{L}(f|c) \propto Z^{-1}(c) \exp\left[-\frac{1}{2\sigma^2} \parallel L(c)f \parallel^2\right] \tag{13}$$

where the operator $L(c) = \Delta + aI - N(c)$ is defined in (9), (10), and the partition function is given by

$$Z(c) = [\det L(c)]^{-1} \tag{14}$$

A circumstance that may appear strange is that here our operator $L(c)$ would be represented by a 4-dimensional array, not a matrix, and the determinant has to be interpreted accordingly. But this is only a superficial difficulty. It is more convenient to avoid the representation and simply write

$$\det L(c) = \prod_i \lambda_i \tag{15}$$

where λ_i, the eigenvalues of $L(c)$, are the focus of our calculation. This is just what we said about the Jacobian in (5).

Consider the eigenvalues λ'_{st} for the operator $\Delta - N(c)$. Note that adding $-N(c)$ to Δ amounts to a small modification of Δ affecting only a minor fraction of the l^4 coefficients defining Δ. It is therefore natural to use a perturbation argument.

If A and B symmetric operators and

$$A' = A + \epsilon B,$$

then an eigenvalue λ of A changes into

$$\lambda' = \lambda + \epsilon(B\psi, \psi) + o(\epsilon), \tag{16}$$

where ψ is the normalized eigenfunction associated with λ. In our case the inner product

$$\epsilon(B\psi, \psi) = -\frac{l}{4} \sum_{\overrightarrow{z,z'}} [\psi(z) - \psi(z')]^2 \tag{17}$$

where the sum is over all pairs $\overrightarrow{z,z'}$, $(z = (x, y)$ and $z' = (x', y'))$, that make up the cutset.

However, there is a complication due to the multiple eigenvalues $\lambda_{st} = \lambda_{ts}$ for $s \neq t$. This forces us to employ a modified form of (16) [cf. Kato (1966), pp 81-83 and 120-121]. For each (s, t) (or(t, s)), the modification of (16) for double eigen-values is given by

$$\lambda'_{st} = \lambda_{st} + \epsilon k_{st}^{(1)} + o(\epsilon) \tag{18}$$

$$\lambda''_{st} = \lambda_{st} + \epsilon k_{st}^{(2)} + o(\epsilon), \tag{19}$$

where $k_{st}^{(1)}$ and $k_{st}^{(2)}$ are the two eigenvalues of the matrix

$$K = \begin{pmatrix} (B\psi_{st}, \psi_{st}) & (B\psi_{st}, \psi_{ts}) \\ (B\psi_{ts}, \psi_{st}) & (B\psi_{ts}, \psi_{ts}) \end{pmatrix} \tag{20}$$

Only the sum $k_{st}^{(1)} + k_{st}^{(2)}$ will be needed later on, hence we only calculate the trace of K.

Notice that a pair $\overrightarrow{z, z'}$ is either horizontal

$$x' = x + 1, \;\; y' = y \tag{21}$$

or vertical

$$x' = x, \;\; y' = y + 1 \tag{22}$$

For large l, by routine calculation we have

$$\begin{aligned} \epsilon(k_{st}^{(1)} + k_{st}^{(2)}) &= \epsilon[(B\psi_{st},\ \psi_{st}) + (B\psi_{ts},\ \psi_{ts})] && (23) \\ &\approx -\frac{4}{l^2}\Bigg(\sin^2\frac{\pi s}{2l} \sum_H \sin^2\frac{\pi t y}{l} \cos^2\frac{\pi s x}{l} && (24) \\ &\quad + \sin^2\frac{\pi t}{2l} \sum_V \sin^2\frac{\pi s x}{l} \cos^2\frac{\pi t y}{l} && (25) \\ &\quad + \sin^2\frac{\pi t}{2l} \sum_H \sin^2\frac{\pi s y}{l} \cos^2\frac{\pi t x}{l} && (26) \end{aligned}$$

$$+\sin^2\frac{\pi s}{2l}\sum_V \sin^2\frac{\pi tx}{l}\cos^2\frac{\pi ty}{l}\Big), \tag{27}$$

$$\tag{28}$$

where $\sum_H$ is over all horizontal pairs $\overrightarrow{z, z'}$ in the cutset and $\sum_V$ over all vertical pairs.

If we assume that the cutset c consists of a finite number of smooth arcs located in the interior of the unit square discretized into an $l \times l$ lattice, then by setting $u = \frac{x}{l}, v = \frac{y}{l}, du \approx \frac{1}{l} dv \approx \frac{1}{l}$, we obtain the curve integral approximation

$$\epsilon(B\psi_{st}, \psi_{st}) \tag{29}$$

$$\approx \quad -\frac{4}{l}\int_c \sin^2\frac{\pi s}{2l}\cdot\sin^2\pi tv\cdot\cos^2\pi su \ du + \sin^2\frac{\pi t}{2l}\cdot\sin^2\pi su\cdot\cos^2\pi tv \ dv.$$

Now let

$$D(\epsilon) = \det L(c) = \prod_{s,t}(\lambda'_{st}+a) = \prod_s(\lambda'_{ss}+a)\cdot\prod_{s<t}(\lambda'_{st}+a)(\lambda''_{st}+a) \tag{30}$$

$$D(O) = \det(\Delta + aI) = \prod_{s,t}(\lambda_{st}+a). \tag{31}$$

Then

$$\log\frac{D(\epsilon)}{D(0)} \approx \sum_s \frac{1}{\lambda_{ss}+a}\epsilon(B\psi_{ss},\psi_{ss}) + \sum_{s<t}\frac{1}{\lambda_{st}+a}\,\epsilon(k^{(1)}_{st}+k^{(2)}_{st})$$

$$= \sum_{s,t}\frac{1}{\lambda_{st}+a}\,\epsilon(B\psi_{st},\psi_{st}).$$

Setting $\xi = \frac{s}{l}, \zeta = \frac{t}{l}$, $d\xi \approx \frac{1}{l}, d\zeta \approx \frac{1}{l}$ and using (18),(19),(20), we obtain the approximation

$$\log \quad \frac{D(\epsilon)}{D(0)} \tag{32}$$

$$\stackrel{\dagger}{\approx} \quad - \quad 4l\int_c \left(\int_0^1 \int_0^1 \frac{\sin^2\frac{\pi\xi}{2}\cdot\sin^2\pi l\zeta\upsilon\cdot\cos^2\pi l\xi u}{1-\frac{1}{2}\cos\pi\xi-\frac{1}{2}\cos\pi\zeta+a}d\xi\,d\zeta\right)du \tag{33}$$

$$\tag{34}$$

$$+ \quad \left(\int_0^1 \int_0^1 \frac{\sin^2\frac{\pi\zeta}{2}\cdot\sin^2\pi l\xi u\cdot\cos^2\pi l\zeta\upsilon}{1-\frac{1}{2}\cos\pi\xi-\frac{1}{2}\cos\pi\zeta+a}d\xi\,d\zeta\right)d\upsilon \tag{35}$$

$$\tag{36}$$

$$\stackrel{\ddagger}{\approx} \quad - \quad \frac{l}{2}\int_c \left(\int_0^1 \int_0^1 \frac{1-\cos\pi\xi}{1-\frac{1}{2}\cos\pi\xi-\frac{1}{2}\cos\pi\zeta+a}d\xi\,d\zeta\right)du \tag{37}$$

$$\tag{38}$$

$$+ \quad \left(\int_0^1 \int_0^1 \frac{1-\cos\pi\zeta}{1-\frac{1}{2}\cos\pi\xi-\frac{1}{2}\cos\pi\zeta+a}d\xi\,d\zeta\right)d\upsilon \tag{39}$$

$$\tag{40}$$

$$= \quad - \quad \frac{l}{2}\left(\int_0^1 \int_0^1 \frac{1-\cos\pi\xi}{1-\frac{1}{2}\cos\pi\xi-\frac{1}{2}\cos\pi\zeta+a}d\xi\,d\zeta\right)\left(\int_c du+d\upsilon\right), \tag{41}$$

$$\tag{42}$$

where ($\dagger$) follows from Fubini Theorem and ($\ddagger$) follows from Riemann-Lebesgue Lemma based on the fact that for large l, the functions $\sin^2\pi l\xi u, \cos^2\pi l\xi u$ oscillate rapidly with the mean value $\frac{1}{2}$ over the interval $0 \leq \xi \leq 1$, as do the functions $\sin^2\pi l\zeta\upsilon, \cos^2\pi l\zeta\upsilon$ over the interval $0 \leq \zeta \leq 1$.

4 Pattern Inference for Cutset Processes

To organize pattern inference we shall, as in several earlier instances, simulate the posterior density by an SDE which requires, as its main part, the computation of the gradient of the logarithm of the posterior density.

The part due to the prior causes no problem, it is dealt with as before. The other part, due to the likelihood, is radically different however, and requires more attention. The inference can deal with several hypothetical cutsets, differring in location, pose, and shape. Then the partiton functions also differ and their computation is neccessary.

Since the partition function $Z(c)$ has been calculated in section 3, we now consider the quadratic form $Q = \| L(c)f \|^2$. We can write

$$\begin{aligned} Q &= ([\Delta + aI - N(c)]f,\ [\Delta + aI - N(c)]f) \\ &= \| (\Delta + aI)f \|^2 - 2((\Delta + aI)f,\ N(c)f) + (N(c)f,\ N(c)f)). \end{aligned} \tag{43}$$

Using the notation $f = G$ and $(\Delta + aI)f = F$, let us calculate $(F,\ N(c)G)$. The third term $(N(c)f,\ N(c)f)$ in (43) can be handled similarly.

Note that

$$(F,\ N(c)G) = \frac{1}{4} \sum_{\overrightarrow{z,z'}} [F(z) - F(z')][G(z) - G(z')], \tag{44}$$

where the sum is over all pairs $\overrightarrow{z,z'}$ that form the cutset.

For large l we get the curve integral approximation

$$(F, N(c)G) \approx \frac{l}{4} \int_c (H,\ dz) \tag{45}$$

with the column vector

$$H = \begin{pmatrix} \dfrac{\partial F}{\partial x} \cdot \dfrac{\partial G}{\partial x} \\ \\ \dfrac{\partial F}{\partial y} \cdot \dfrac{\partial G}{\partial y} \end{pmatrix} \tag{46}$$

assuming sufficient smoothness for F, G and c thought of as given in the continuum $[0,1]^2$.

We can write, using a continuous formulation,

$$z(t) = s^0 + s^1 c^0(t), \quad 0 \le t \le n \tag{47}$$

with

$$c^0(t) = \int_0^t g^0(s)ds, \quad g^0(s) = g_i^0 \ \text{ for } \ i \le s < i+1 \tag{48}$$

where $s^0 = (s_x^0, s_y^0)^t$ is a translation, s^1 is a scalar meaning scale change. For the moment we have made all the individual rotation angles $\phi_i = 0$; we can later on include the ϕ_i.

We need the second derivatives $\frac{\partial^2 z}{\partial t \partial \theta}$ where θ stands for one of the group element parameters. For $\theta = s_x^0$ and s_y^0 this second derivative is zero. For $\theta = s^1$ it is equal to $\frac{dc^0(t)}{dt}$ which equals g_i^0 for $i \le t < i+1$. Hence,

$$\frac{\partial}{\partial \theta}(F, N(c)G) \approx \frac{l}{4}\int_{t=0}^{n} \left(H, \frac{\partial^2 z}{\partial t \partial \theta}\right) dt + \frac{l}{4}\int_{t=0}^{n} \left(\frac{\partial H(z(t))}{\partial \theta}, \frac{\partial z}{\partial t}\right) dt \tag{49}$$

with

$$\frac{\partial H(z(t))}{\partial \theta} = \frac{\partial H}{\partial x}\frac{\partial x(t)}{\partial \theta} + \frac{\partial H}{\partial y}\frac{\partial y(t)}{\partial \theta} = \left(\nabla H, \frac{\partial z(t)}{\partial \theta}\right) \tag{50}$$

where ∇H denotes the 2×2 matrix

$$\nabla H = \begin{pmatrix} \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial x} \cdot \frac{\partial G}{\partial x}\right) & \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial y} \cdot \frac{\partial G}{\partial y}\right) \\ \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial x} \cdot \frac{\partial G}{\partial x}\right) & \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial y} \cdot \frac{\partial G}{\partial y}\right) \end{pmatrix} \tag{51}$$

This completes the calculation of the quantities needed for inference.

5 Acknowledgment.

This paper has gained from discussions with Professor Chuanshu Ji.

6 Bibliography.

U. Grenander (1993): General Pattern Theory, Oxford University Press

U. Grenander, M.I. Miller (1998): Computational Anatomy: An Emerging

Discipline, Quarterly of Applied Mathematics, LVI, No. 4,617-694

T. Kato (1966): Perturbation Theory for Linear Operators, Springer-Verlag

A. Lanterman, U. Grenander, M.I. Miller (2000): Bayesian Segmentation Via Asymptotic Partition Functions, IEEE Trans. on Pattern Analysis and Machine Intelligence, 22,4,337-347, April, 2000.