An Invariant Variational Principle for Model-Based Interpolation of High Dimensional Clustered Data

R. C. Venkatesan
Systems Research Corporation
Aundh, Pune 411007, INDIA
ravi@systemsresearchcorp.com

Abstract

A self-consistent scheme based on the calculus of infinitesimal transformations that describes model based interpolation (MBI hereafter) of high dimensional data, constrained to lie on a nonlinear manifold, is expressed as a dynamical system. The variational formulation derived for a single cluster is extended to the case of finite mixture models (FMM hereafter). The suggested formulation is shown to be qualitatively dissimilar properties, and exhibits greater computational efficiency, as compared with a scheme derived using "classical" variational calculus.

1 Introduction

Nonlinear interpolation is one aspect of interpolation in high dimension space, which finds immense utility in important applications as diverse as image processing [1], audio-visual speech recognition [2], video email [3], and low bandwidth teleconferencing [4].

Prior attempts to model the process of nonlinear interpolation have utilized radial basis functions [5], locally linear models [1, 2], and more recently, the seminal work of Saul and Jordan (the S-J model hereafter) using "classical" variational principles [6]. The present work bears a fundamental similarity with the S-J model in that the pdf induces the metric (the arc length increment between any two points) and geometry of the space. The term nonlinear interpolation (which is used interchangeably with MBI), specifically refers to interpolation of data, which are constrained to lying on a nonlinear manifold (surface).

The present model is characterized by probability density functions (pdf's hereafter), which approximate data are modeled, rather than individual data points. Specifically, the data is taken to exist as clusters (regions of high density inhabiting a high dimensional space), that constitute a FMM [7] whose model parameters are fixed. Such a semi-parametric representation of the data is made possible by evoking the fact that in statistical pattern recognition, a gray-level image having n*m *pi(x)*cture *ele*ments may be described by a single point lying in a n*m dimensional space.

The present analysis limits itself to the case of Gaussian pdf's, which provide excellent locally linear approximations to nonlinear manifolds, for the case of images. Further, the full covariance matrices associated with Gaussian pdf's tend to simplify linear algebra operations.

The rationale for the present study is to investigate into the possibility of applying the scheme derived herein to the field of digital compression. There are essentially two ways in which data points may be transmitted. First, all the points could be transmitted in sequence.

An alternate approach is to encode the images into a system of differential equations, which are transmitted along with the initial conditions of the trajectory. The set of differential equations that describe the evolution of the system could then be integrated to compute the points along the trajectory. This could result in great savings in communication (or compression) if the differential equation describes a large number of points along the trajectory.

The present analysis extends the S-J model by suggesting a formulation based on the calculus of infinitesimal transformations [8], instead of employing "classical" variational methods. This results in a set of differential relations known as conservation laws being derived from Noether's theorem. The preservation of invariances has been found to play an important part in statistical learning and vision systems [9]. The use of Noether's theorem leads to a formulation, which is qualitatively different, and, possesses greater computational tractable than the S-J model.

2 The Variational Model

The present analysis commences with the fact that the data is already represented in the form of a FMM, whose model parameters are known. The clusters are obtained as a consequence of applying a locally linear approximation to the nonlinear surface. In this manner, a projection of the data points of the image feature space (that comprise a low dimensional nonlinear manifold), onto locally linear subspaces that approximate the nonlinear manifold, is performed.

The prototype (maxima of the pdf) centers are located on the data points by means of k-means clustering. Next, a PCA is performed on a specified number of nearest neighbors of each prototype, with the intention of calculating the dimension of the locally linear manifold.

Further refinement is performed using the EM algorithm. The present analysis is based on the premise that a continuous vector **x**(t) traces a smooth multi-dimensional curve (known as a trajectory) to points, segmented by a discrete variable z. In the present model, both **x**(t), and z evolve temporally.

The discrete (hidden) variable, z, assigns labels to different points in space traversed by the trajectory generated by **x**(t), at consecutive time intervals. The trajectory generated by x(t) is segmented by mapping it into a sequence of labels traced by z. This sequence takes the form of a piecewise constant function in time, z(t). ***The task of learning involves the determination of the locations of the points where the discrete variable changes values***. ***A-priori knowledge of the model parameters classifies the present study as being a case of supervised learning.***

FMM interpolation may be considered as being a chain of intra-cluster interpolations linked together at points where the discrete variable changes values. FMM interpolation comprises of two separate processes – intra-cluster interpolation (the *continuous model* - characterized by $P[\mathbf{x}(t)|z = z(t)]$ and inter-cluster interpolation (the *segmentation model* - characterized by $P[z|x(t) = constant / fixed]$), respectively.

FMM interpolation is essentially an iterative process, and in general does not readily lend itself to computational implementation. In the present model, the procedure is greatly simplified by expressing the *continuous model* as a variational principle, using the MLE theory. Extending the continuous model to include all possible clusters that comprise the FMM, the evaluation of the optimal interpolation trajectory passing through the "k" clusters is obtained.

The Lagrangian functional associated with the *continuous model* (for a single cluster) is generically expressed as $L[(\mathbf{x}(\mathrm{t}), (\dot{\mathbf{x}}(\mathrm{t})]$, and that associated with the overall process of FMM interpolation is expressed as $L_{z(t)}[(\mathbf{x}(\mathrm{t}), \dot{\mathbf{x}}(\mathrm{t})]$. By definition, the *continuous model* depends upon an *a-priori* segmentation of the path length. Hence, we first describe the process that is responsible for the segmentation. The expression for $L_{z(t)}[\bullet]$ is obtained by extending the variational for the continuous model to the case of FMM's. From the theory of FMM, we have a super-population G of Gaussian pdf's which comprises a finite mixture of "k" populations $G_1,...,G_k$, in some known mixing proportion $\pi_1,...,\pi_k$, respectively. The analysis commences by deriving a variational formulation for the continuous trajectory for the case of intra-cluster interpolation, assuming an *a-priori* temporal segmentation of the trajectory generated by **x**(t). Let:

$$P(z) = \sum_{z=1}^{k} \pi_z = 1, \tag{1}$$

where P(z) is a discrete pdf over "k" possible outcomes. The pdf of a random vector in G may be most generally expressed as:

$$P(\mathbf{x}, \mathrm{z}) = \sum_{z=1}^{k} \pi_z P(\mathbf{x}|z). \tag{2}$$

In our model, z(t) is a piecewise constant function of time, which temporally segments the inter-cluster path in parameterized space. Specifically, the task of z(t) is to ascribe different variationals (Lagrangians) to different regions of space (clusters), between time intervals $t + \delta t$. Here, z(t) is numerically evaluated by calculating the most probable segmentation:

$$z(t) = \underset{\mathrm{z}}{\mathrm{argmax}} \left[L_{z(t)}[(\mathbf{x}(\mathrm{t}), (\dot{\mathbf{x}}(\mathrm{t})] \right], \tag{3}$$

for "k" possible values of z, and then choosing the maximum value of $L_{z(t)}[\bullet]$. This is performed using a modified EM algorithm. Eq. (3) comprises a hard clustering assignment. Clustering nomenclature defines "hard" assignments as those which give rise to definite distribution, while "soft" assignments are those which give rise to a probability distribution. The utility of (3) is to determine the boundary crossing that divides the interpolation path into points assigned to adjacent clusters, respectively. Specifically, for a given intra-cluster whose end points are separated by a finite amount of time t_{i-1} and t_i, $z(t) = z_i$.

The value of z(t) is discontinuous across the boundaries separating contiguous clusters. As was with the case with the *continuous model*, the present operation also has the effect of minimizing the action functional corresponding to

$L_{z(t)}[\bullet]$, thus maximizing the likelihood (minimizing the negative log-likelihood) of the interpolation path.. The trajectory x(t) within a specific cluster is determined as follows – consider a set of multi-dimensional points $X = \{\mathbf{x}^{(1)},...,\mathbf{x}^{|N|}\}$, and P(**x**) denote the pdf of the parameterized model associated with these points. The present analysis derives an action for intra-cluster interpolation, satisfying two separate and simultaneously applicable constraints. The first of the constraints is that the path length linking the data points passes through regions of space, which are assigned a high probability. The second constraint was that the path length be minimized in order to avoid deformations.

The present theory maximizes the probabilities of data points along a single interpolation path. An alternate strategy to formulate the variational principle is based on the conjecture that maximizing the probability of each point on the path does not equate to maximizing the probability of the path. In such a scenario, it would be necessary to assign appropriate distribution over paths, and then to find the most likely path within that distribution. The present analysis does not treat the relative merits/de-merits between the two approaches to modeling the problem of FMM interpolation. This will be the task of future work.

Assuming an *a-priori* segmentation of the path length of the continuous model, each segment is assigned a weight of the form:

$$\phi(t) = \left[\frac{P(\mathbf{x})}{P(x^*)}e^{-l}\right]^{D[x(t)+x(t+dt)]}, \quad (4)$$

In (4), $D[\bullet] = \left[\sum_{\alpha\beta} g_{\alpha\beta}(q)dx_\alpha dx_\beta\right]^{1/2}$ is the Euclidean metric having the metric tensor $g_{\alpha\beta}(\mathbf{x}) = -\frac{\partial^2[\ln P(\mathbf{x}(t))]}{\partial x_\alpha \partial x_\beta}$.

Herein a simple Euclidean metric is employed. However, the MBI theory admits more complicated metrics, such as the Riemananian metric. Here P(x*) represents the unique maxima of the pdf, and is referred to as the *prototype*. A constant 'l" whose role is similar to that of line tension in the theory of splines is included. From (1), it is evident that $0 < \phi(t) \le 1$. The similarity between the metric tensor and the Fisher information matrix may be noted. Taking the continuous limit, and specifying that $e^{-S[\bullet]} = \prod_t \phi(t)$, the action functional is obtained as the negative log-likelihood as:

$$S[\mathbf{x}(t)] = \int\left\{-\ln\left[\frac{P(\mathbf{x}(t))}{P(x^*)}\right]+l\right\}\left[\sum_{\alpha\beta} g_{\varepsilon\beta}dx_\alpha dx_\beta\right]^{1/2} dt. \quad (5)$$

The weights assigned to each segmented path length ensure that minimization of the cost function maximizes the number of data points interpolated within a given cluster. . It should be noted that the action functionals are essentially representatives (placeholders) of the pdf's. A possible drawback that exists within the present theory is that the minimization of the action functional might tend towards a local, rather than a global minima, owing to the concave nature of the cost function.

To extend the above intra-cluster interpolation to the case FMM interpolation, we ascribe an inter-cluster line tension of the form :

$$l_z = l - \ln \pi_z. \quad (6)$$

The inter-cluster action functional is expressed as:

$$\begin{aligned} &S[\mathbf{x}(t), z(t)] \\ &= \int\left\{-\ln\left[\sum_{z=1}^{k}\frac{\pi_z P(\mathbf{x}|z)}{P(x_z^*)}\right]+l\right\}\left[\sum_{\alpha\beta} g_{\alpha\beta}dx_\alpha dx_\beta\right]^{1/2} dt. \quad (7) \\ &= \int L_{z(t)}[\mathbf{x}(\mathbf{t}), \dot{\mathbf{x}}(t)]dt \end{aligned}$$

Here, $P(x_z^*)$ is the prototype of the z^th cluster.

2.1 Noether's Theorem and its Utility in MBI

Consider the set of one-parameter infinitesimal transformations:

$$\begin{aligned} \bar{t}(t, x_i, \dot{x}_i; a) &= \Psi(t, x_i, \dot{x}_i; a) = t(\bullet) + \delta a \xi(\bullet) \\ \bar{x}_i(t, x_i, \dot{x}_i; a) &= \Phi_i(t, x_i, \dot{x}_i; a) = x_i(\bullet) + \delta a \eta_i(\bullet), \quad (8) \\ \dot{\bar{x}}_i(t, x_i, \dot{x}_i; a) &= \Phi_{i,t}(t, x_i, \dot{x}_i; a) = \dot{x}_i(\bullet) + \delta a \dot{\eta}_i(\bullet). \end{aligned}$$

Here, "a" is the group parameter, ξ and η_i are the coordinate functions of the independent and dependent variables, respectively. The one-parameter group generator of infinitesimal transformations and its first extension are given in indicial form by:

$$U = \xi(t, x_i)\frac{\partial}{\partial t} + \eta_i(t, x_i)\frac{\partial}{\partial x_i}., U^{(1)} = U + \dot{\eta}_i\frac{\partial}{\partial \dot{x}_i}. \quad (9)$$

Here, $\dot{\eta}_i = \frac{D\eta_i}{Dt} - \frac{dx_i}{dt}\frac{D\xi}{Dt}$, and $\frac{D}{Dt} = \frac{\partial}{\partial t} + \frac{\partial x_i}{\partial t}\frac{\partial}{\partial x_i} + \ldots$

Given Δ to be the non-simultaneous variation (as opposed to the "classical" variational δ), defined in indicial form as:

$$\Delta \int_{x_i(t_1)}^{x(t_2)} L[\bullet])dt = \int_{\bar{x}_i(t_1)}^{\bar{x}(t_2)} L(\bar{x}_i(t), \dot{\bar{x}}_i(t))dt - \int_{x_i(t_1)}^{x(t_2)} L(x_i(t), \dot{x}_i(t))dt \,, \quad (10)$$

Noether's theorem is defined by the invariance of the Hamilton action integral with respect to the infinitesimal transformations defined in (8), is expressed as:

$$\Delta S[\bullet] = \Delta \int_{\mathrm{x}_i(t_1)}^{\mathrm{x}_2(t_2)} \mathrm{L}[\bullet]\mathrm{dt} = 0 = \int_{\mathrm{x}_i(t_1)}^{\mathrm{x}_2(t_2)} \delta a \frac{DF[\bullet]}{Dt} dt \quad , \qquad (11)$$

where, " $\mathrm{F}[\bullet]$ " is an arbitrary gauge function. To obtain the values of the coordinate functions, we evaluate Noether's fundamental identity:

$$U^{(1)} L[\bullet] + L[\bullet]\frac{D\xi}{Dt} = 0 \,. \qquad (12)$$

The conservation laws may be obtained from the expression:

$$\frac{D}{Dt}\left[F[\bullet] - \left(\frac{\partial L[\bullet]}{\partial \dot{x}_i}\right)\eta_i(\bullet)\right] = 0. \qquad (13)$$

Here, (13) has been derived using prior knowledge that $\xi = 0$***, (see Section 3), and is contingent to the condition that the E-Le's be satisfied***. The allowance for time deformation in Noether's theorem (non-simultaneous variations) does not, in general, permit commutation between the differential operator D and the non-simultaneous variation Δ. On the other hand, "classical" variational calculus (simultaneous variations), permits commutation between the differential operator and the "classical" variation operator δ. When $\xi = 0$, Δ coincides with δ. This dovetails with the requirement that the variational be invariant to re-parameterization of the time ($t \to f(t)$, where $f(t)$ is a monotonic increasing function). From (13) we observe that conservation laws obtained from Noether's theorem do not possess additional terms corresponding to boundary conditions. Specifically, the terms corresponding to the re-adjustment of the boundary crossings between contiguous clusters within the framework of the S-J model are absorbed into the expression for the conservation laws.

This distinction vis-à-vis "classical" variational calculus has the effect of profoundly affecting the continuous model, the evaluation of the boundary crossings between contiguous clusters, and, the optimization described by (3) for the case of FMM interpolation. Details highlighting the distinction are provided in Section 4 of the present paper.

3 Noether's Theorem for Multi-Dimensional Gaussian Pdf's

Consider the expression for the multi-dimensional Gaussian pdf :

$$P(\mathbf{x}) = \frac{|M|^{1/2}}{(2\pi)^{N/2}} \exp\left[-(\mathbf{x}-\mu)^T M(\mathbf{x}-\mu)\right], \qquad (14)$$

where $M = \begin{bmatrix} m_{11} & 0 \\ 0 & m_{22} \end{bmatrix}$ is the inverse covariance matrix, N is the dimensionality, and μ is the vector of means. Here, $(\mathbf{x}-\boldsymbol{\mu})^T M(\mathbf{x}-\boldsymbol{\mu})$ is the Mahalanobis distance. Assuming that the mean coincides with the origin, we obtain the action for the uni-modal, multi-dimensional (2-D) *keeping z(t) at a fixed value*:

$$S[\mathbf{x}(t)] = \int \left\{\frac{1}{2}[\mathbf{x}^T M\mathbf{x}] + l\right\}\left[\dot{\mathbf{x}}^T \; M\,\dot{\mathbf{x}}\right]^{1/2} . \qquad (15)$$

We obtain the value of the one-parameter group generator of infinitesimal transformations (the variational symmetries) to be:

$$U = x_2 \frac{\partial}{\partial x_1} - \frac{m_{11}}{m_{22}} x_1 \frac{\partial}{\partial x_2} . \qquad (16)$$

Equation (20) signifies a rotational symmetry, which corresponds to the conservation of angular momentum. The resulting E-Le's are expressed as:

$$\left\{\frac{1}{2}\left[\mathbf{x}^T \mathbf{M}\mathbf{x}\right] + l\right\}\ddot{\mathbf{x}} + (\mathbf{x}^T M\,\dot{\mathbf{x}})\dot{\mathbf{x}} = \mathbf{x}\,, \qquad (17)$$

The law of conservation of angular momentum, obtained from Noether's theorem, is expressed as:

$$\left\{\frac{1}{2}\left[\mathbf{x}^T M\mathbf{x}\right] + l\right\}\ddot{\mathbf{x}} + (\mathbf{x}^T M\,\dot{\mathbf{x}})\dot{\mathbf{x}} = 0\,. \qquad (18)$$

Here, (16)-(18) have been derived employing the fact that in the case of multi-dimensional Gaussian pdf, the inverse covariance matrix (which is symmetric) is the metric. Since the variational formulation for multi-dimensional Gaussian pdf's is invariant to re-parameterization of time, and time does not appear in an explicit manner in the form of the action functional, the condition $\dot{\mathbf{x}}^T M \dot{\mathbf{x}} = 1$ is employed in the derivation of (17) and (18), and is enforced in the numerical integration of (18). *Equation (18) assumes immense utility in the area of image interpolation because "real images", which exist in very high dimensional spaces, may be simply interpolated by solving a system of "N" coupled ODE's.*

The ODE's described by (18) are numerically integrated as a BVP using a modified version of the COLSYS routine [10], implemented in Fortran 95. The inverse covariance matrix is taken to be the identity matrix ($m_{11} = m_{22} = 1$). Sample results for three different values of "l" are depicted in Figure 1. For real images inhabiting very high dimensional spaces, calculation of the symmetry properties is somewhat involved, owing to the large number of dependent variables. The difficulties are readily mitigated by utilizing symbolic packages such as MATHEMATICA® [11].

4 Extension to the case of FMM interpolation

It is important to note the distinction between the model parameters (the location of the clusters and their boundaries-that are assumed to be known a-priori), and the boundary crossings of the interpolating trajectory (that have to be evaluated). The FMM interpolation is viewed within the context of a state-space model that is solved using a two-step approach – segmentation and interpolation. To numerically evaluate (3) for time t in the range $[0, \tau]$, we employ a dynamic programming technique along the lines of the forward-backward Viterbi algorithm [12]. Keeping **x**(t) at a fixed value at a given iteration level, the temporal axis is discretized in increments of Δt. The derivative terms present in $L_{z(t)}[\bullet]$ are approximated by their finite-difference analogs. Thus, we obtain:

$$\begin{aligned} z(t) &= \arg\max_z \left[L_{z(t)}[\mathbf{x}(t), \dot{\mathbf{x}}(t)] \right] \\ &= \arg\max_z \left[\sum_{z=1}^{k} \frac{1}{2} \left\{ \mathbf{x}(t)^T M \mathbf{x}(t) + l - \ln \pi_z \right\} \times \right. \\ &\left. \left[\dot{\mathbf{x}}(t)^T M \dot{\mathbf{x}}(t) \right]^{1/2} \left\{ = M^{1/2} \frac{x(t+\Delta t) - x(t)}{\Delta t} \right\} \right] \end{aligned} \qquad (19)$$

The above process may be summarized as follows - the value of z that maximizes the functional $L_{z(t=0)}[\bullet]$ is first obtained. The time is incremented in steps of Δt, till another value of z is found that that maximizes $L_{z(t)}[\bullet]$. The procedure is repeated until an enforced END value of $\tau = 1$ is reached. At this juncture, we obtain a piecewise constant function z(t) that defines the boundary crossing between contiguous clusters, the exact value of t for which $z(t) = \underset{z}{\operatorname{argmax}} \left[L_{z(t)}[\bullet] \right]$ changes value.

Given two adjacent clusters G_i, and G_{i+1} drawn from a super-population G, (3)/(19) utilizes the model parameters to obtain the point that divides the interpolation path between G_i, and G_{i+1} with end points $\mathbf{x}_i(t)$ and $\mathbf{x}_{i+1}(t)$. Next, the *continuous model* is solved for each cluster by evaluating (18) between $\mathbf{x}_i(t) \to \mathbf{x}(t_c)$ and $\mathbf{x}(t_c) \to \mathbf{x}_{i+1}(t)$. The process is repeated in an iterative manner till convergence is reached. In the S-J model, the condition that the *continuous model* satisfies the E-Le's is enforced contingent to the fact that the additional terms in the first order optimality condition for the action functional, representing the boundary conditions, are already satisfied. The sub-paths are obtained by integrating the continuous model between $\mathbf{x}_i(t) \to \mathbf{x}(t_c)$ and $\mathbf{x}(t_c) \to \mathbf{x}_{i+1}(t)$. Here too, the boundary conditions are provided by the hard clustering assignment. To satisfy the first order optimality condition of the action functional and make the boundary crossings of the *segmentation model* and the boundary conditions of the *continuous model* concomitant with one another at a given iteration level, a gradient descent operation solving $\frac{\delta S}{\delta x} = \left. \frac{\partial L}{\partial \dot{\mathbf{x}}(t)} \right|_{t_{c-}}^{t_{c+}}$ is performed. Here, t_{c-} and t_{c+} signify points lying on either side of the interface dividing regions defined by two discrete states. The intra-cluster interpolation is re-performed yielding a new trajectory through $\mathbf{x}(t_c)$. This process ensures a smooth transition of the interpolation trajectory between contiguous clusters. *This process has been referred to as the "re-stitching" phase in the S-J model. The boundary crossing correction in the S-J model could turn out to be a prohibitively expensive computational procedure, especially when one bears in mind the high dimensions in which "real life" images are embedded into.*

To provide a numerical example, the case of FMM interpolation for six Gaussian clusters is demonstrated. The means are distributed in a manner so as to lie on a circle. The inverse covariance matrix is taken to be the identity matrix. As is evident from Figure 2, as the iteration progresses, the FMM interpolation approaches the circle of means. The case of linear interpolation is used to provide for an initial condition. Since the means correspond to the prototypes of the clusters, and represent regions of maximum data distribution, the fidelity of the FMM interpolation, within the context of the present model may be easily observed.

5 Discussions, Conclusions, and Future Work

The framework for FMM interpolation using Noether's theorem has been established. The ramifications in feature space of the invariance preserving model describing the dynamic evolution of an optimal interpolating trajectory in parameterized space for the case of FMM's, is an important issue which warrants detailed study. When employing the present scheme for image data transmission, it is important to study the nature of the dynamical system within the context of chaotic paths and limit cycles. Initial studies concerning the above topics have been performed, which will be reported elsewhere for lack of space herein.

Future work will also accomplish the parameterization of variational models for MBI, based on a set of invariances that have been specified *a-priori*, within the framework of Takens method for the inverse problem in the calculus of infinitesimal transformations [13]. This theory requires that the Frechet derivatives of the E-Le's equal zero (the Helmholtz condition). The present framework does not satisfy this condition. Commencing with a source equation and a set of pre-specified invariances, the weights assigned to the segmented path lengths of the continuous model are treated as being arbitrary functions of the pdf and the line tension. This provides the prospect to derive new variational principles depending upon the types of invariances to be preserved. Such an approach is already in development, and also seeks to specify an explicit relationship between the line tension "l" (which in the present theory is a free and unspecified parameter), and the induced metric.

Acknowledgements

This work has been performed under the auspices of STR&DF contract MSR-DCSM 03-1999-201982-215710. This work was partially carried out at Project Robot-Vision (ROBOTVIS), INRIA (Sophia-Antipolis). Gratitude is felt towards L. K. Saul, A. H. Salden, O. Faugeras, C. Bregler, and S. Omohundro for helpful discussions.

References

[1] C. Bregler and S. Omohundro, "Nonlinear Image Interpolation Using Manifold Learning" in Advances on Neural Information Processing Systems, edited by G. Tesauro, D. Tburetzky, & T. Leen, MIT Press, Cambridge MA (1995).

[2] C. Bregler and S. Omohundro, "Nonlinear Manifold Learning for Visual Speech Recognition", Proc. of the 50th Intl. Conf. on Computer Vision, edited by E.Grimson, pp 494–499, IEEE Computer Society Press, Los Alamitos, CA (1995).

[3] T. Ezzat, "Example Based Analysis and Synthesis for Images of Faces", MIT, EECS M. S. Thesis (1996).

[4] D. Beymer, A. Shashua, and T. Poggio, "Example Based Image Analysis and Synthesis", MIT AI Memo 1161 (1997).

[5] T. Poggio and F. Girosi, "Networks for Approximation and Learning", Proc. IEEE, 78, 9, (1990).

[6] L. K. Saul and M. I. Jordan, "A Variational Principle for Model Based Interpolation", .Advances in Neural Information Processing Systems 9,edited by M. C. Mozer, M. I. Jordan, & T. Petsche, Cambridge MA: MIT Press (1997).

[7] G. McLachlan and K. Basford, "Mixture Models: Inference and Applications to Clustering", Marcel Dekker (1988).

[8] P. J. Olver, "Application of Lie Groups to Differential Equations", Springer-Verlag, New York (1996).

[9] P. Y. Simard, Y. LeCun, and J. Denker, "Efficient Pattern Recognition Using a New Transformation Distance", Advances in Neural Information Processing Systems, 5, 50-58, Morgan Kauffman, San Mateo, CA (1993).

[10] U. Ascher, J. Christiansen and R. D. Russell, "Collocation Software for Boundary-Value ODE's", ACM Trans. Math Software, 7, 209-222 (1981).

[11] A. J. Viterbi, "Error Bounds for Convolution Codes and an Asymptotically Optimal Decoding Algorithm", IEEE Trans. Inf. Theory, 13, 260-269 (1967).

[12] Gerd Baumann, "Symmetry Analysis of Differential Equations with Mathematica", Springer-Verlag (TELOS), New York (2000)

[13] F. Takens, "Symmetries, Conservation Laws, and Variational Principles", Lecture Notes in Math 597, pp581-603, Springer-Verlag, Heidelberg (1977).

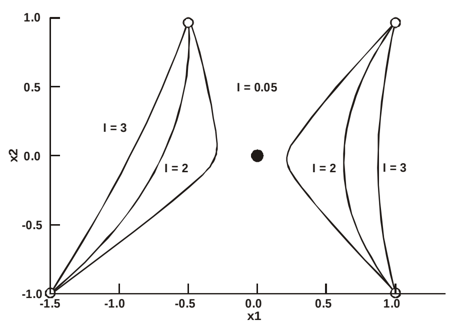

Figure 1: Simulation Results for Intra-Cluster Interpolation Using Conversation Laws.

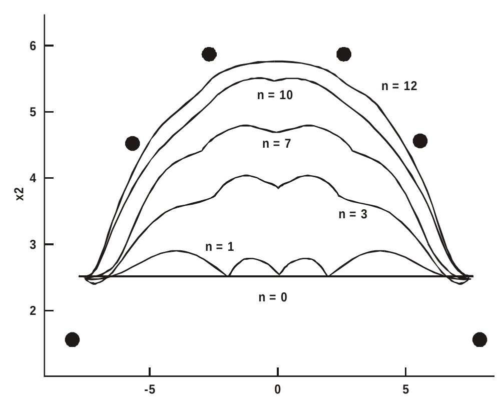

Figure 2: Simulation Results Finite Mixture Model for Interpolation Using Conversation Laws.