**Rutgers Computer Science Technical Report RU-DCS-TR589**
**Dec. 2005, revised May 2006**

# Impact of Dynamics on Subspace Embedding and Tracking of Sequences

by

Kooksang Moon and Vladimir Pavlović
Rutgers University
Piscataway, NJ 08854
ksmoon@paul.rutgers.edu, vladimir@cs.rutgers.edu

# ABSTRACT

In this paper we study the role of dynamics in dimensionality reduction problems applied to sequences. We propose a new family of *marginal auto-regressive* (MAR) models that describe the space of all stable auto-regressive sequences, regardless of their specific dynamics. We apply the MAR class of models as sequence priors in probabilistic sequence subspace embedding problems. In particular, we consider a Gaussian process latent variable approach to dimensionality reduction and show that the use of MAR priors may lead to better estimates of sequence subspaces than the ones obtained by traditional non-sequential priors. We then propose a learning method for estimating nonlinear dynamic system (NDS) models that utilizes the new MAR priors. The utility of the proposed methods is demonstrated on several synthetic datasets as well as on the task of tracking 3D articulated figures in monocular image sequences.

# 1 Introduction

Dimensionality reduction / subspace embedding methods such as Principal Components Analysis (PCA), Multidimensional Scaling (MDS), Gaussian Process Latent Variable Models (GPLVM) [9] and others, play an important role in many data modeling tasks by selecting and inferring those features that lead to an intrinsic representation of the data. As such, they have attracted significant attention in computer vision where they have been used to represent intrinsic spaces of shape, appearance, and motion. However, it is common that subspace projection methods applied in different contexts do not leverage inherent properties of those contexts. For instance, subspace projection methods used in human figure tracking [4, 12, 16, 17] often do not fully exploit the dynamic nature of the data. As a result, the selected subspaces sometimes do not exhibit temporal smoothness or periodic characteristics of the motion they model. Even if the dynamics are used, the methods employed are sometimes not theoretically sound and are disjoint from the subspace selection phase.

In this paper we present a new approach to subspace embedding of sequential data that explicitly accounts for their dynamic nature. We first model the space of sequences using a novel Marginal Auto-Regressive (MAR) formalism. A MAR model describes the space of sequences generated from all possible AR models. In the limit case MAR describes all *stable* AR models. As such, the MAR model is weakly-parametric and can be used as a prior for an arbitrary sequence, without knowing the typical AR parameters such as the state transition matrix. The embedding model is then defined using a probabilistic Gaussian Process Latent Variable (GPLVM) framework [9] with MAR as its prior. A GPLVM framework is particularly well suited for this task because of its probabilistic generative interpretation. The new hybrid GPLVM and MAR framework results in a general model of the space of all *nonlinear dynamic systems* (NDS). Because of this it has the potential to theoretically soundly model nonlinear embeddings of a large family of sequences.

The paper is organized as follows. We first define the family of MAR models and study some properties of the space of sequences modeled by MAR. Next, we show that MAR and GPLVM result in a model of the space of all NDS sequences and discuss its properties. The utility of the new framework is examined through a set of experiments with synthetic and real data. In particular, we apply the new framework to modeling and tracking of the 3D human figure motion from a sequence of monocular images.

# 2 Marginal Auto-Regressive Model

## 2.1 Definition

Consider sequence $X$ of length $T$ of $N$-dimensional real-valued vectors $x_t = [x_{t,0} x_{t,1} ... x_{t,N-1}] \in \Re^{1 \times N}$. Suppose sequence $X$ is generated by the 1st order AR model $AR(A)$:

$$x_t = x_{t-1} A + w_t, \; t = 0, ..., T-1, \tag{1}$$

where $A$ is a specific $N \times N$ state transition matrix and $w_t$ is a white iid Gaussian noise with unit precision, $w_t \sim \mathcal{N}(0, I)$. Assume that, without loss of generality, the initial condition $x_{-1}$ has normal multivariate distribution with zero mean and unit precision $\alpha_i$: $x_{-1} \sim N(0, \alpha_i^{-1} I)$.

We adopt a convenient representation of sequence $X$ as a $T \times N$ matrix $X = [x_0' x_1' ... x_{T-1}']'$ whose rows are the vector samples from the sequence. Using this notation (1) can be written as

$$X = X_\Delta A + W, \tag{2}$$

where $W = [w_0' w_1' ... w_{T-1}']'$ and $X_\Delta$ is a *shifted/delayed* version of $X$, $X_\Delta = [x_{-1}' x_0' ... x_{T-2}']'$ . Given the state transition matrix $A$ and the initial condition, the AR sequence samples have the joint density function

$$P(X|A, x_{-1}) = (2\pi)^{NT/2} \exp\left\{ -\frac{1}{2} tr \left\{ (X - X_\Delta A)(X - X_\Delta A)' \right\} \right\}. \tag{3}$$

The density in (3) describes the distribution of samples in a $T$-long sequence for a particular instance of the state transition matrix $A$. However, we are interested in the distribution of all AR sequences, regardless of the value of $A$. In other words, we are interested in the marginal distribution of AR sequences, over all possible parameters $A$.

Assume that all elements $a_{ij}$ of $A$ are iid Gaussian with zero mean and precision $\alpha$, $a_{ij} \sim \mathcal{N}(0, \alpha^{-1})$. Under this assumption, one can show that the *marginal* distribution of the AR model becomes

$$P(X|x_{-1}, \alpha) = \int_A P(X|A, x_{-1}) P(A|\alpha) dA = (2\pi)^{NT/2} |K_{xx}(X,X)|^{-1} exp\left\{ \frac{1}{2} tr\{ K_{xx}(X,X)^{-1} X X' \} \right\} \tag{4}$$

where

$$K_{xx}(X,X) = X_\Delta X_\Delta' + \alpha^{-1} I. \tag{5}$$

We call this density the *Marginal AR* or MAR density. $\alpha$ is the hyperparameter of this class of models, $MAR(\alpha)$. Intuitively, (4) favors those samples in $\mathcal{X}$ that do not change significantly from $t$ to $t+1$ and $t-1$.

MAR density models the distribution of all (AR) sequences of length $T$ in the space $\mathcal{X} = \Re^{T \times N}$. Note that while the error process of an AR model has Gaussian distribution, the MAR density is not Gaussian. We illustrate this in Fig. 1. The figure shows pdf values for four different densities: MAR, periodic MAR (see Sec. 2.2), AR(2),and a circular Gaussian, in the space of length-2 scalar-valued sequences $[x_0 x_1]'$. In all four cases we assume zero-mean , unit precision Gaussian distribution of the initial condition. All models have the mode at $(0,0)$. However, the variance of the AR model is elliptical, with axes determined by the state transition matrix $A$. The MAR models define non-Gaussian distributions with no circular symmetry and with directional bias. This property of MAR densities is important when viewed in the context of sequence subspace embeddings, which we discuss in Sec. 3.

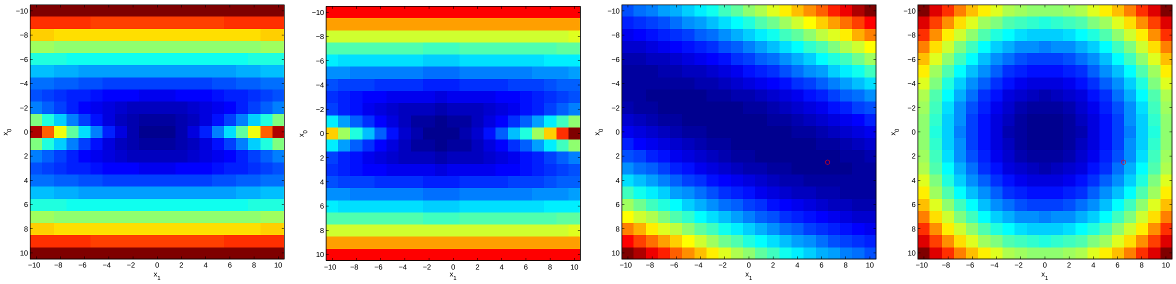

Figure 1: Distribution of length-2 sequences of 1D samples under MAR, periodic MAR, AR, and independent Gaussian models.

## 2.2 Higher-Order Dynamics

The above definition of MAR models can be easily extended to families of arbitrary $D$-th order AR sequences. In that case the state matrix $A$ is replaced by an $ND \times N$ matrix $A = [A_1' A_2' ... A_D']'$ and $X_\Delta$ by $[X_\Delta X_{1\Delta} ... X_{D\Delta}]$. Hence, a $MAR(\alpha, D)$ model describes a general space of all $D$-th order AR sequences. Using this formulation one can also model specific classes of dynamic models. For instance, a class of all periodic models can be formed by setting $A = [A_1' \quad -I]'$, where $I$ is an identity matrix.

## 2.3 Nonlinear Dynamics

In (1) and (4) we assumed linear families of dynamic systems. One can generalize this approach to nonlinear dynamics of the form $x_t = g(x_{t-1}|\zeta)A$, where $g(\cdot|\zeta)$ is a nonlinear mapping to an $L$-dimensional subspace and $A$ is a $L \times N$ linear mapping. In that case $K_{xx}$ becomes a nonlinear kernel using justification similar to e.g. [9]. While nonlinear kernels often have potential benefits, such as robustness, they also preclude closed-form solutions of linear models. In our preliminary experiment we have not observed significant differences between MAR and nonlinear MAR.

## 2.4 Justification of MAR Models

The choice of the prior distribution of AR model's state transition matrix leads to the MAR density in (4). One may wonder, however, if the choice of iid $\mathcal{N}(0, \alpha^{-1})$ results in a physically meaningful space of sequences. We suggest that, indeed, such choice may be justified.

Namely, Girko's circular law [5] states that if $\frac{1}{N}A$ is a random $N \times N$ matrix with $\mathcal{N}(0, 1)$ iid entries, then in the limit case of large $N$ (¿20) all real and complex eigenvalues of $A$ are *uniformly distributed on the unit disk*. For small $N$, the distribution shows a concentration along the real line. Consequently, the resulting space of sequences described by the MAR model is that of *all stable AR systems*.

# 3 Nonlinear Dynamic System Models

In this section we develop a Nonlinear Dynamic System view of the sequence subspace reconstruction problem that relies on the MAR representation of the previous section. In particular, we use the MAR model to describe the structure of the subspace of sequences to which the extrinsic representation will be mapped using a Gaussian Process latent variable model of [9].

## 3.1 Definition

Let $Y$ be an extrinsic or measurement sequence of duration $T$ of $M$-dimensional samples. Define $Y$ as the $T \times M$ matrix representation of this sequence, similar to the definition in Sec. 2, $Y = [y_0' y_1' ... y_{T-1}']'$. We assume that $Y$ is a result of the process $X$ in a lower-dimensional MAR subspace $\mathcal{X}$, defined by a nonlinear generative or forward mapping

$$Y = f(X|\theta)C + V. \tag{6}$$

$f(\cdot)$ is a nonlinear $\Re^M \rightarrow \Re^L$ mapping, $C$ is a linear $L \times N$ mapping, and $V$ is a zero-mean unit variance Gaussian noise.

To recover the intrinsic sequence $X$ in the embedded space from sequence $Y$ it is convenient not to focus, at first, on the recovery of the specific mapping $C$. Hence, we consider the family of mappings where $C$ is a stochastic matrix whose elements are iid $c_{ij} \sim \mathcal{N}(0, \beta^{-1})$. Marginalizing over all possible mappings $C$ yields a marginal Gaussian Process [19] mapping:

$$\begin{aligned} P(Y|X,\beta,\theta) &= \int_C P(Y|X,C,\theta)P(C|\beta)dC \\ &= (2\pi)^{NT/2}|K_{yx}(X,X)|^{-1}exp\left\{\frac{1}{2}tr\{K_{yx}(X,X)^{-1}YY'\}\right\} \end{aligned} \tag{7}$$

where

$$K_{yx}(X,X) = f(X|\theta)f(X|\theta)' + \beta^{-1}I. \tag{8}$$

Notice that in this formulation the $X \rightarrow Y$ mapping depends on the inner product $\langle f(X), f(X) \rangle$. The knowledge on the actual mapping $f$ is not necessary; a mapping is uniquely defined by specifying a positive-definite kernel $K_{yx}(X, X|\theta)$ with entries $K_{yx}(i,j) = k(x_i, x_j)$ parameterized by the hyperparameter $\theta$. A variety of linear and non-linear kernels (RBF, square exponential, various robust kernels) can be used as $K_{yx}$. Hence, our likelihood model is a nonlinear Gaussian process model,as suggested by [9].

In this manner we have constructed a marginal Nonlinear Linear Dynamic System (MNDS) model that describes the joint distribution of all measurement and all intrinsic sequences in a $\mathcal{Y} \times \mathcal{X}$ space:

$$P(X,Y|\alpha,\beta,\theta) = P(X|\alpha)P(Y|X,\beta,\theta). \tag{9}$$

The MNDS model has a MAR prior $P(X|\alpha)$, and a Gaussian process likelihood $P(Y|X,\beta,\theta)$. Thus it places the intrinsic sequences $X$ in the space of all AR sequences. Given an intrinsic sequence $X$, the measurement sequence $Y$ is zero-mean normally distributed with the variance determined by the nonlinear kernel $K_{yx}$ and $X$.

## 3.2 Inference

Given a sequence of measurements $Y$ one would like to infer its subspace representation $X$ in the MAR space, without needing to first determine a particular family of AR models $AR(A)$, nor the mapping $C$. (9) shows that this task can be, in principle, achieved using the Bayes rule $P(X|Y, \alpha, \beta, \theta) \sim P(X|\alpha)P(Y|X, \beta, \theta)$.

However, this posterior is non-Gaussian because of the nonlinear mapping $f$ and the MAR prior. One can instead attempt to estimate the mode $X^*$

$$X^* = \arg\max_X \{\log P(X|\alpha) + \log P(Y|X, \beta, \theta)\} \tag{10}$$

using nonlinear optimization such as the Scaled Conjugate Gradient in [9].

To effectively use a gradient-based approach, one needs to obtain expressions for gradients of the log-likelihood and the log-MAR prior. Note that the expressions for MAR gradients are more complex than those of e.g. GP due to a linear dependency between $X$ and $X_\Delta$.

## 3.3 Learning

MNDS space of sequences is parameterized using a set of hyperparameters $(\alpha, \beta, \theta)$ and the choice of the nonlinear kernel $K_{yx}$. Given a set of sequences $\{Y^{(i)}\}, i = 1, .., S$ the learning task can be formulated as the ML/MAP estimation problem

$$(\alpha^*, \beta^*, \theta^*)|_{K_{yx}} = \arg\max_{\alpha,\beta,\theta} \prod_{i=1}^{S} P(Y^{(i)}|\alpha, \beta, \theta). \tag{11}$$

One can use a generalized EM algorithm to obtained the ML parameter estimates recursively from two fixed-point equations:

**E-step**:
$X^{(i)*} = \arg\max_X P(Y, X^{(i)}|\alpha^*, \beta^*, \theta^*)$

**M-step**:
$(\alpha^*, \beta^*, \theta^*) = \arg\max_{(\beta,\alpha,\theta)} \prod_{i=1}^{K} P(Y^{(i)}, X^{(i)*}|\alpha, \beta, \theta)$

## 3.4 Learning of Explicit NDS Model

Inference and learning in MNDS models result in the embedding of the measurement sequence $Y$ into the space of all NDS/AR models. Given $Y$, the embedded sequences $X$ estimated in Sec. 3.3 and MNDS parameters $\beta, \alpha, \theta$, the explicit AR model can be easily reconstructed using the ML estimation on sequence $X$, e.g. :

$$A^* = (X'_\Delta X_\Delta)^{-1} X'_\Delta X. \tag{12}$$

Because the embedding was defined as a GP, the likelihood function $P(y_t|x_t, \beta, \theta)$ follows a well-known result from GP theory

$$y_t|x_t \sim \mathcal{N}(\mu, \sigma^2 I)$$

$$\mu = Y' K_{yx}(X, X)^{-1} K_{yx}(X, x_t)$$

$$\sigma^2 = K_{yx}(x_t, x_t) - K_{yx}(X, x_t)' K_{yx}(X, X)^{-1} K_{yx}(X, x_t).$$

The two components fully define the explicit NDS.

In summary, a complete sequence modeling algorithm consist of the following set of steps:

```
Input  : Measurement sequence $Y$ and kernel family $K_{yx}$
Output : $NDS(A, \beta, \theta)$
1) Learn subspace embedding $MNDS(\alpha, \beta, \theta)$ model of training sequences $Y$, Sec. 3.3.
2) Learn explicit subspace and projection model $NDS(A, \beta, \theta)$ of $Y$, Sec. 3.4.
```

Algorithm 1: NDS learning.

## 3.5 Inference in Explicit NDS Model

The choice of the nonlinear kernel $K_{yx}$ results in a nonlinear dynamic system model of training sequences $Y$. The learned model can then be used to infer subspace projections of a new sequence from the same family. Because of the nonlinearity of the embedding, one cannot apply the linear forward-backward or Kalman filtering/smoothing inference. Rather, it is necessary to use nonlinear inference methods such as (I)EKF or particle filtering/smoothing.

It is interesting to note that one can often use a relatively simple sequential nonlinear optimization in place of the above two inference methods:

$$x_t^* = \arg\max_{x_t} P(y_t|x_t, \beta^*, \theta^*) P(x_t|x_{t-1}^*, A^*). \tag{13}$$

Such sequential optimization yields local modes of the true posterior $P(X|Y)$. While one would expect such approximation to be valid in situations with few ambiguities in the measurement space and models learned from representative training data, our experiments show the method to be robust across a set of situations. However, dynamics seem to play a crucial role in the inference process.

## 3.6 Example

We illustrate the concept of MNDS on a simple synthetic example. Consider the AR model $AR(2)$ from Sec. 2. Sequence $X$ generated by the model is projected to the space $\mathcal{Y} = \Re^{2\times 3}$ using a linear conditional Gaussian model $\mathcal{N}(XC, I)$. Fig. 2 shows negative likelihood over the space $\mathcal{X}$ of the MNDS, a marginal model (GP) with independent Gaussian priors, a GP with the exact$AR(2)$ prior, and a full LDS with exact parameters. All likelihoods are computed for the fixed $Y$. Note that the GP with Gaussian prior assumes no temporal structure in the data. This example shows that, as expected, the optimal subspace estimates of the MNDS model fall closer to the “true” LDS estimates than those of the the non-sequential model. This property holds in general. Fig. 3 shows the distribution of optimal negative log likelihood scores, computed at corresponding $X^*$, of the four models over a 10000 sample of $Y$ sequences generated from the true LDS model. Again, one notices that MNDS has a lower mean and mode than the non-sequential model, GP+Gauss, indicating MNDS's better fit to the data. This suggests that MNDS may result in better subspace embeddings than the traditional GP model with independent Gaussian priors.

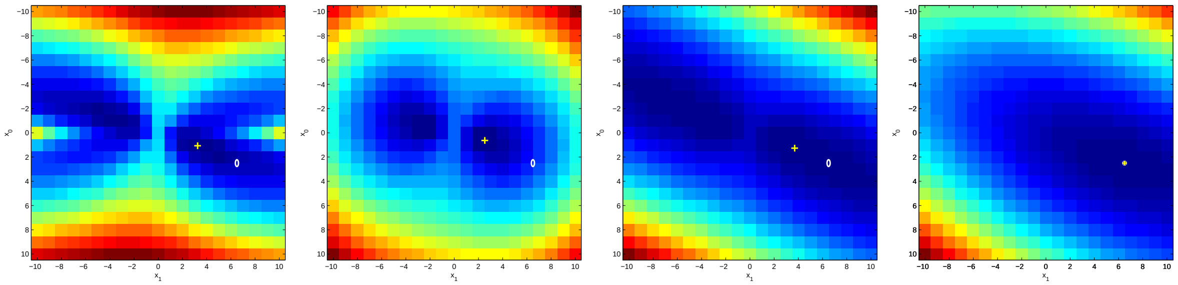

Figure 2: Negative log-likelihood of length-2 sequences of 1D samples under MNDS, GP with independent Gaussian priors, GP with exact AR prior and LDS with the true process parameters. “o” mark represents the optimal estimate $X^*$ inferred from the true LDS model. “o” shows optimal estimates derived using the three marginal models.

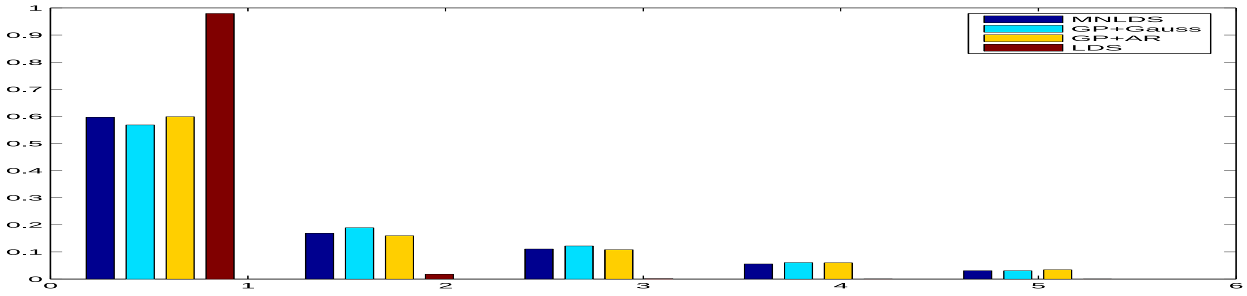

Figure 3: Histogram of optimal negative log-likelihood scores for MNDS, a model with a Gaussian prior, and the LDS with the true parameters.

# 4 Human Motion Modeling using MNDS

In this section we consider an application of MNDS to modeling of the human motion from sequences of video images. Specifically, we assume that one wants to recover two important aspects of human motion: (1) 3D posture of the human figure in each image and (2) an intrinsic representation of the motion.

We propose the following model in this context. Given a sequence of features $z_t$ computed from monocular images (such as the silhouette-based alt moments, orientation histograms etc. ), the mapping into the 3D pose space represented by joint angles $y_t$ is given by a Gaussian process model $P(Y|Z, \theta_{yz})$ with a parametric kernel $K_{yz}(z_t, z_t|\theta_{yz})$. An NDS is used to model the space $\mathcal{Y} \times \mathcal{X}$ of poses and intrinsic motions $P(X, Y|A, \beta, \theta_{yx})$.

The joint *conditional* model of the pose sequence $Y$ and intrinsic motion $X$, given the sequence of image features $Z$ is approximated by

$$P(X, Y|Z, A, \beta, \theta_{yz}, \theta_{yx}) \approx P(Y|Z, \theta_{yz})P(X, Y|A, \beta, \theta_{yx}). \tag{14}$$

The reason for this approximation is practical—modeling $P(Y|Z)$ rather than $P(Z|Y)$ yielded better results and allowed a fully GP-based framework.

## 4.1 Learning

In the training phase, both the image features $Z$ and the corresponding poses $Y$ are known. Hence, the learning of GP and NDS models becomes decoupled and can be accomplished using the NDS learning formalism presented in the previous section and a standard GP learning approach [19].

```
Input  : Image sequence $Z$ and joint angle sequence $Y$
Output : Human motion model.
1) Learn Gaussian Process model $P(Y|Z, \theta_{yz})$ using e.g. [19].
2) Learn NDS model $P(X, Y|A, \beta, \theta_{yx})$ as described in Sec. 3.
```

Algorithm 2: Human motion model learning.

## 4.2 Inference and Tracking

Once the models are learned they can be used for tracking of the human figure in video. Because both NDS and GP are nonlinear mappings, estimating current pose (distribution) $y_t$ given a previous pose and intrinsic motion space estimates $P(x_{t-1}, y_{t-1}|Z_{0..t})$ will involve nonlinear optimization or linearizion, as suggested in Sec. 3.5. In particular, optimal point estimates $x_t^*$ and $y_t^*$ are the result of the following nonlinear optimization problem:

$$(x_t^*, y_t^*) = \arg\max_{x_t, y_t} P(x_t|x_{t-1}, A)P(y_t|x_t, \beta, \theta_{yx})P(y_t|z_t, \theta_{yz}). \tag{15}$$

The point estimation approach is particularly suited for a particle-based tracker. Unlike some traditional approaches that only consider the pose space representation, tracking in the low dimensional intrinsic space has the potential to avoid problems associated with sampling in high-dimensional spaces.

A sketch of the human motion tracking algorithm is shown below.

```
Input  : Image $z_t$, prior point estimates $(w_{t-1}^{(i)}, x_{t-1}^{(i)}, y_{t-1}^{(i)})|Z_{0..t-1}, i = 1, ..., S$ and Human
         motion model (GP +NDS).
Output : Current pose/intrinsic state estimates $(w_t^{(i)}, x_t^{(i)}, y_t^{(i)})|Z_{0..t}$
1) Choose the initialize estimates $x_t^{(i)}, y_t^{(i)}$ among the training data $(X, Y, Z)$ using nearest
neighbor matching in $Z$ space.
2) Find optimal estimates $(x_t^{(i)}, y_t^{(i)})$ using nonlinear optimization in (15).
3) Find point weights $w_t^{(i)} \sim P(x_t^{(i)}|x_{t-1}, A)P(y_t^{(i)}|x_t^{(i)}, \beta, \theta_{yx})P(y_t^{(i)}|z_t, \theta_{yz})$.
```

Algorithm 3: Human motion tracking.

We apply this algorithm to a set of tracking problems described in Sec. 6.2.

# 5 Related Work

Manifold learning approaches to motion modeling have attracted significant interest in the last several years. Brand proposed nonlinear manifold learning that maps sequences of the input to paths of the learned manifold [3]. Rosales and Sclaroff [10] proposed the Specialized Mapping Architecture (SMA) that utilizes forward mapping for the pose estimation task. Agarwal and Triggs [1] directly learned a mapping from image measurement to 3D pose using Relevance Vector Machine (RVM).

However, it is often advantageous to consider a subspace of e.g. the joint angles space that contains a compact representation of the actual figure motion. Non-linear manifold embedding of the training data in low dimensional spaces using isometric feature mapping (Isomap), Local linear (LLE) and spectral embedding [15, 11, 2, 18], have shown success in recent approaches [4, 12]. While these techniques provide point-based embeddings implicitly modeling the nonlinear manifold through exemplars, they lack a fully probabilistic interpretation of the embedding process.

The GPLVM, a Gaussian Processes [19] model, produces a continuous mapping between the latent space and the high dimensional data in a probabilistic manner [9]. Grochow et al. [6] use a SGPLVM to model inverse kinematics for interactive computer animation. Tian et al. [16] use a GPLVM to estimate the 2D upper body pose from the 2D silhouette features. More recently, Urtasun et al. [17] exploit the SGPLVM for 3D people tracking. However these approaches utilize simple temporal constraints in the pose space that often introduce "dimensionality curse" to nonlinear tracking methods such as particle filters. Moreover, such methods fail to explicitly consider motion dynamics during the embedding process. Our work addresses both of these issues through the use of MNLDS models.

# 6 Experiments

## 6.1 Synthetic Data

In our first experiment we examine the utility of MAR priors in a subspace selection problem. A 2nd order AR model is used to generate sequences in a $\Re^{T\times 2}$ space; the sequences are then mapped to a higher dimensional nonlinear measurement space.

$$\begin{aligned} x_t &= A_1 x_{t-1} + A_2 x_{t-2} + w_t \\ y_{t,1} &= x_{t,1}\cos(x_{t,1}) + v_{t,1} \\ y_{t,1} &= x_{t,1}\sin(x_{t,1}) + v_{t,2} \\ y_{t,3} &= x_{t,2} + v_{t,3}. \end{aligned}$$

An example of the measurement sequence, a periodic curve on the Swiss-roll surface, is depicted in Fig. 4.

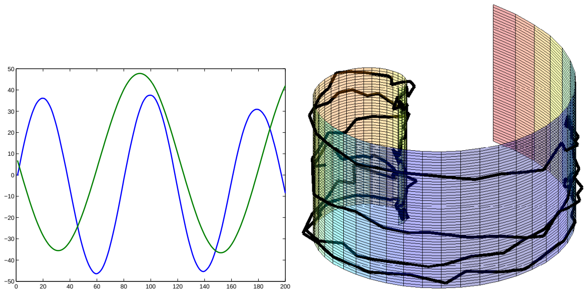

Figure 4: A periodic sequence in the intrinsic subspace and the measured sequence on the Swiss-roll surface.

We apply two different methods to recover the intrinsic sequence subspace: MNDS with an RBF kernel and a GPLVM with the same kernel and independent Gaussian priors. Estimated embedded sequences are shown in Fig. 5. The intrinsic motion sequence inferred by the MNDS model more closely resembles the "true" sequence in Fig. 4. Note that one dimension (blue/dark) is reflected about the horizontal axis, because the embeddings are unique up to an arbitrary rotation. These results confirm that proper dynamic priors may have crucial role in learning of embedded sequence subspaces. We study the role of dynamics in tracking in the following section.

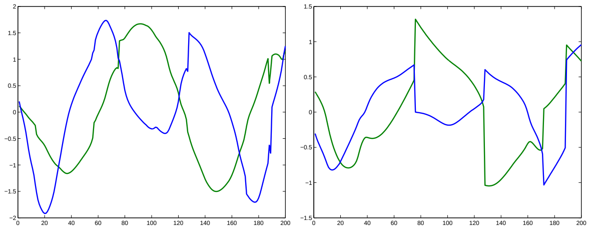

Figure 5: Recovered embedded sequences:MNDS and GPLVM with iid Gaussian priors.

### 6.2 Human Motion Data

We conducted experiments using a database of motion capture data for a 59 d.o.f body model from CMU Graphics Lab Motion Capture Database [7]. Similar to [1, 16] we utilize synthetic images as our training data. Our database consists of seven walking sequences of around 2000 frames total. The data was generated using the software (3D model and Maya binaries) generously provided by the authors of [14, 13]. We train our GP and NDS models with one sequence of 250 frames and test on the remaining sequences. In our experiments, we exclude 15 joint angles that exhibit small movement during walking (e.g. clavicle and fingers joint) and use the remaining 44 joints. Our choice of image features are the silhouette-based Alt moments used in [16, 10]. The scale and translational invariance of Alt moments makes them suitable to a motion modeling task with little or no image-plane rotation.

A portion of the learned latent space is presented in Fig. 6 with a few corresponding silhouette images.

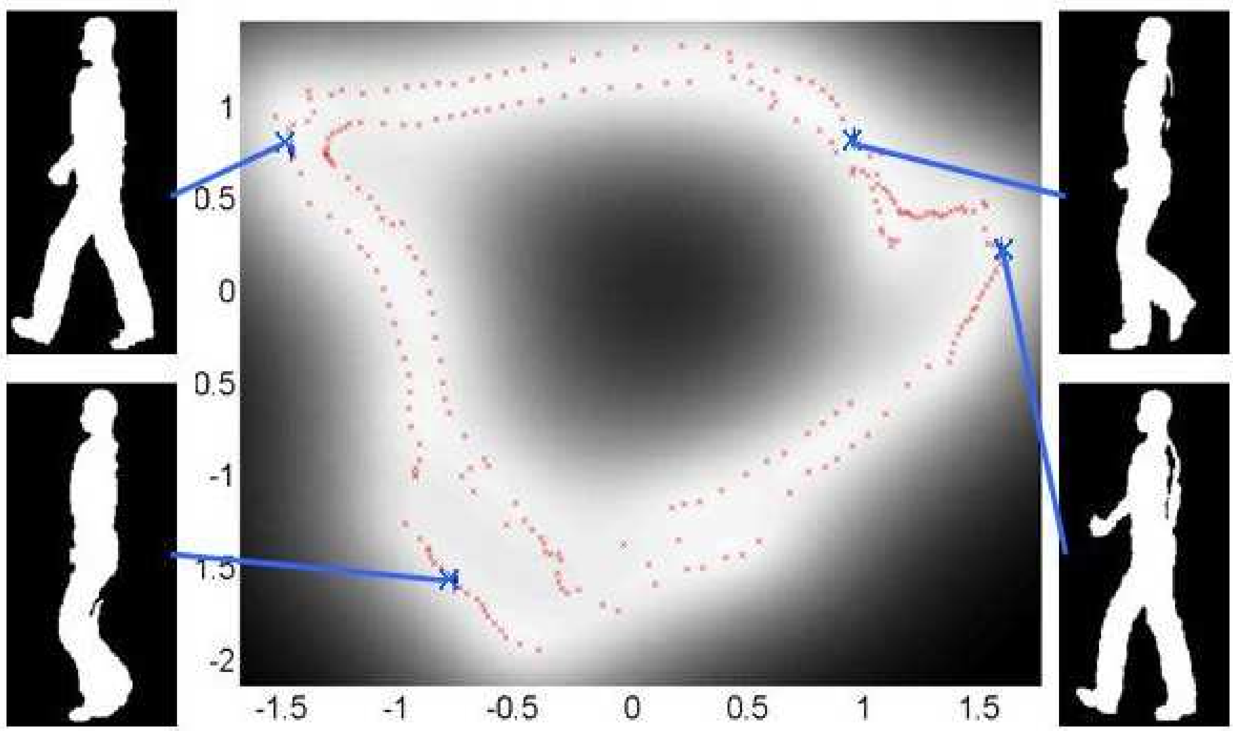

Figure 6: The learned 2-D latent space with of one walking sequence.

In the model learning phase we utilize the approach proposed in Sec. 3. Once the model is learned, we apply the tracking/inference approach in Sec. 4 to infer motion states and poses from sequences of silhouette images. Fig. 7 depicts a sequence of estimated poses. The initial estimates for gradient search are determined by the nearest neighborhood matching in the Alt moments space alone. To evaluate our MNDS model, we estimate the same input sequence with the original GPLVM tracking in [16]. Although the silhouette features are informative for human pose estimation, they are also prone to ambiguities such as the left/right side changes. Without proper dynamics modeling, the original GPLVM fails to estimate the correct poses because of this ambiguity.

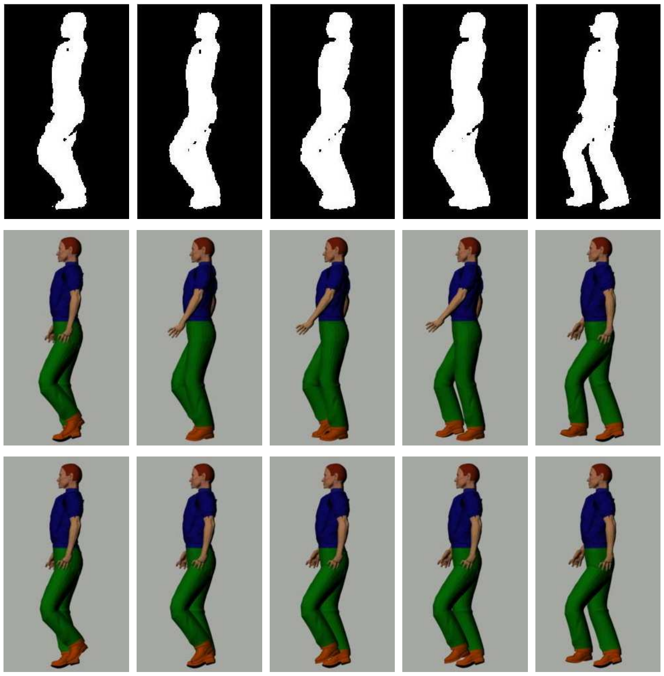

Figure 7: First row: Input image silhouettes. Remaining rows show reconstructed poses. Second row—GPLVM model. Third row—MNDS model.

The accuracy of our tracking method is evaluated using the mean RMS error between the true and the estimated joint angles [1], $D(y, y') = \frac{1}{44}\sum_{i=1}^{44} |(y_i - y_i')mod \pm 180^o|$. Fig. 8 displays the mean RMS errors over the 44 joint angles, estimated using three different models. The testing sequence consists of 320 frames. The mean error for MNDS model is in range $3^o \sim 6^o$. The inversion of right and left legs causes significant errors in the original GPLVM model. Introduction of simple dynamics in the pose space similar to [17] was not sufficient to rectify the "static" GPLVM problem.

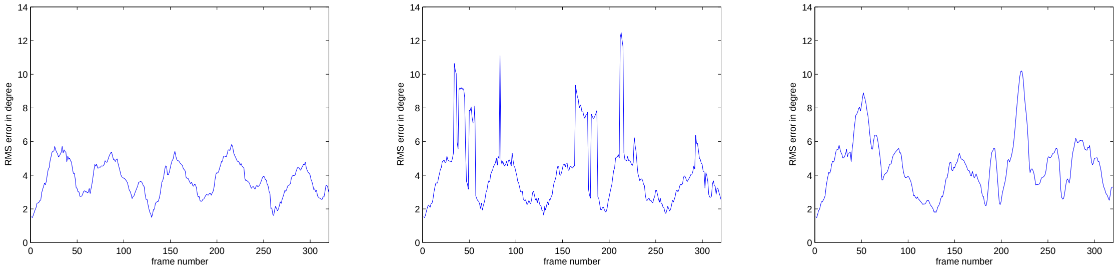

Figure 8: Mean angular pose RMS errors. Left: MNDS model. Middle: original GPLVM model. Right: simple dynamics in the pose space.

Fig. 9 shows examples of trajectories in the embedded space corresponding to the pose estimates. The points inferred from our MNDS model follow the path that is defined by the MAR model, making them temporally consistent. The other two methods produced less-than-smooth embeddings.

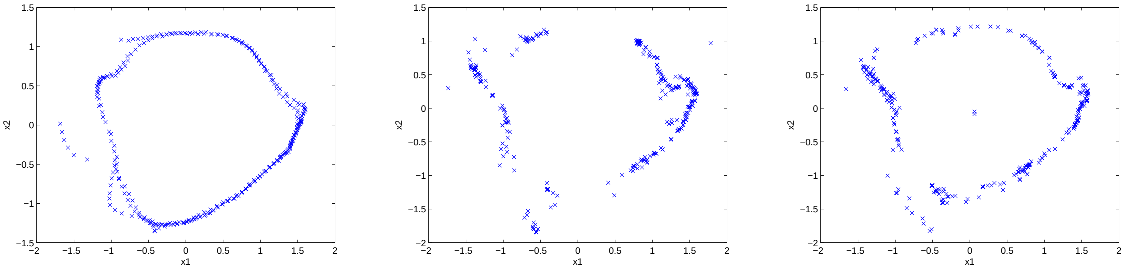

Figure 9: Left: 2D latent space estimates using MNDS. Middle: latent space estimates using the original GPLVM. Right: latent space estimates with simple dynamics in the pose space.

We applied the algorithm to tracking of various real monocular image sequences of the human motion. The data used in these experiments was the sideview sequence in CMU mobo database made publicly available under the HumanID project [8]. Fig. 10 shows one example of our tracking result. This testing sequence consists of 340 frames. Because a slight mismatch in motion dynamics between the training and the test sequences, reconstructed poses are not geometrically perfect. However the overall result sequence depicts a plausible walking motion that agrees with the observed images.

It is also interesting to note that in a number of tracking experiments it was sufficient to carry a very small number of particles ($\sim 1$) in the point-based tracker of Alg.3. In most cases all particles clustered in a small portion of the motion subspace $\mathcal{X}$, even in ambiguous situations induced by silhouette-based features. This indicates that the presence of dynamics had an important role in disambiguating statically similar poses.

## 7 Conclusions

We proposed a novel method for embedding of sequences into subspaces of dynamic models. In particular, we propose a family of marginal AR (MAR) subspaces that describe all stable AR models. We show that a generative nonlinear dynamic system (NDS) can then be learned from a hybrid of Gaussian (latent) process models and MAR priors, a marginal NDS (MNDS). As a consequence, learning of NDS models and state estimation/tracking can be formulated in this new context. Several synthetic examples demonstrate the potential utility of the NDS framework and display its advantages over traditional static methods in dynamic domains. We also test the proposed approach on the problem of the 3D human figure tracking in sequences of monocular images. Our preliminary results indicate that dynamically constructed embeddings using NDS can resolve ambiguities during tracking that may plague static as well as less principled dynamic approaches.

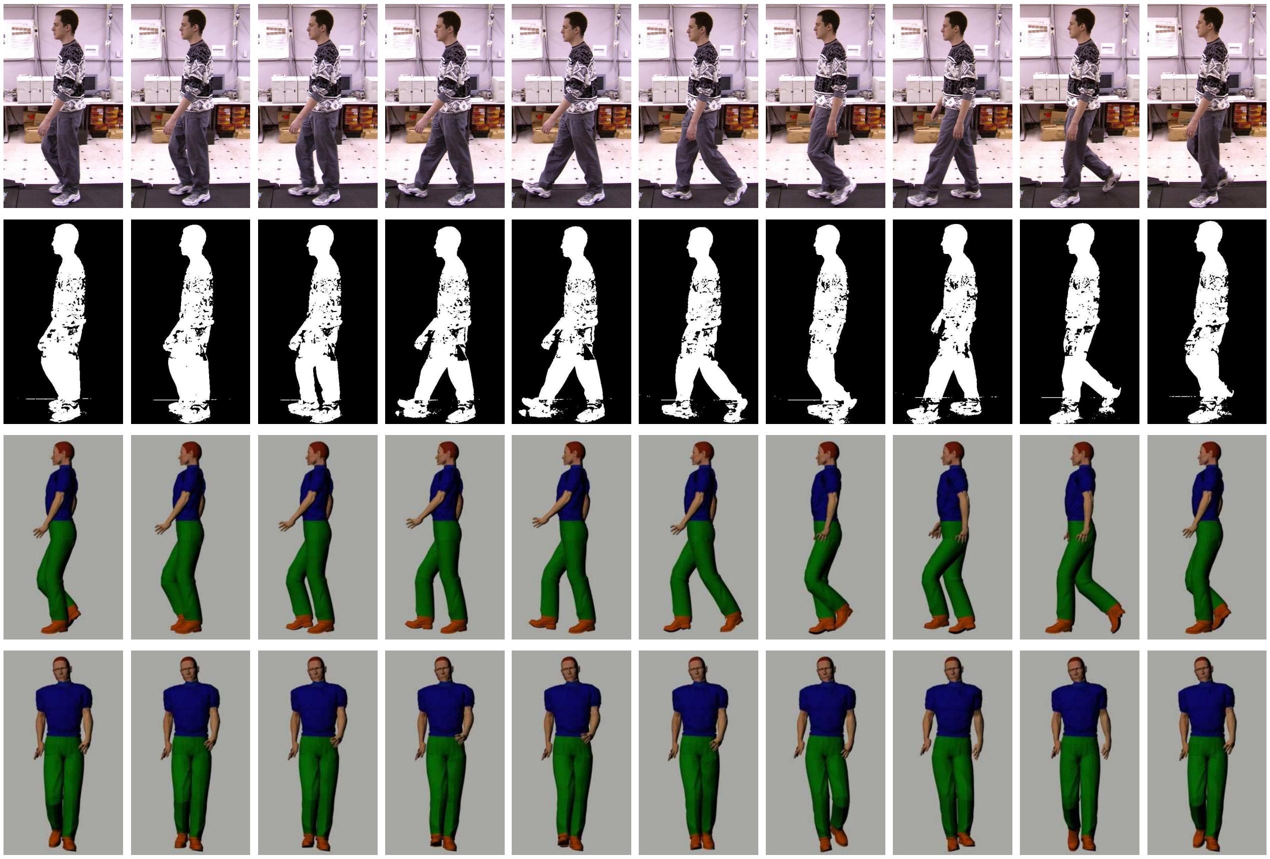

Figure 10: First row: Input real walking images. Second row: Image silhouettes achieved by background subtraction. Third row: Side view of the reconstructed pose. Forth row: Front view of the reconstructed pose.

In our future work we plan to extend the set of evaluations and gather more insight into theoretical and computational properties of MNDS with linear and nonlinear MARs. In particular, our on-going experiments address the posterior multimodality in the embedded spaces, an issue relevant to point-based trackers. We also plan to extend the NDS formalism to collections of dynamic models using the switching dynamics approaches as a way of modeling a general and diverse family of temporal processes.

## 8 Appendix: MAR Gradient

Log likelihood of MAR model is, using (4) and leaving out the constant term,

$$L = \frac{N}{2}\log|K_{xx}| + \frac{1}{2}tr\left\{K_{xx}^{-1}XX'\right\} \tag{16}$$

with $K_{xx} = K_{xx}(X,X)$ defined in (5). Gradient of $L$ with respect to $X$ is

$$\frac{\partial L}{\partial X} = \frac{\partial X_\Delta}{\partial X}\frac{\partial L}{\partial K_{xx}}\frac{\partial K_{xx}}{\partial X_\Delta} + \left.\frac{\partial L}{\partial X}\right|_{X_\Delta}. \tag{17}$$

$X_\Delta$ can be written as a linear operator on $X$,

$$X_\Delta = \Delta \cdot X, \quad \Delta = \begin{bmatrix} 0_{(T-1)\times 1} & I_{(T-1)\times(T-1)} \\ 0 & 0_{1\times(T-1)} \end{bmatrix}, \tag{18}$$

where $0$ and $I$ denote zero vectors and identity matrices of sizes specified in the subscripts. It is now easily follows that

$$\frac{\partial L}{\partial X} = \Delta'\left(NK_{xx}^{-1} - K_{xx}^{-1}XX'K_{xx}^{-1}\right)\Delta \cdot X + K_{xx}^{-1}X. \tag{19}$$

## References

[1] A. Agarwal and B. Triggs. 3d human pose from silhouettes by relevance vector regression. In *CVPR*, pages II 882–888, 2004. 9, 11, 12

[2] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6):1373–1396, 2003. 9

[3] M. Brand. Shadow puppetry. In *CVPR*, volume II, pages 1237–1244, 1999. 9

[4] A. Elgammal and C.-S. Lee. Inferring 3d body pose from silhouettes using activity manifold learning. In *CVPR*, volume 2, pages 681–688, 2004. 1, 9

[5] V. L. Girko. Circural law. *Theory Probab. Appl.*, 29:694–706, 1984. 3

[6] K. Grochow, S. L. Martin, A. Hertzmann, and Z. Popovic. Style-based inverse kinematics. *ACM Trans. Graph.*, 23(3):522–531, 2004. 9

[7] http://mocap.cs.cmu.edu/. 11

[8] http://www.hid.ri.cmu.edu/Hid/databases.html. 13

[9] N. D. Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *NIPS*. 2004. 1, 3, 4, 5, 9

[10] R. Rosales and S. Sclaroff. Specialized mappings and the estimation of human body pose from a single image. In *Workshop on Human Motion*, pages 19–24, 2000. 9, 11

[11] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, December 2000. 9

[12] C. Sminchisescu and A. Jepson. Generative modeling for continuous non-linearly embedded visual inference. In *ICML '04: Proceedings of the twenty-first international conference on Machine learning*, page 96, New York, NY, USA, 2004. ACM Press. 1, 9

[13] C. Sminchisescu, A. Kanaujia, Z. Li, and D. N. Metaxas. Conditional visual tracking in kernel space. *NIPS*, 2005. 11

[14] C. Sminchisescu, A. Kanaujia, Z. Li, and D. N. Metaxas. Discriminative density propagation for 3d human motion estimation. *CVPR*, pages 390–397, 2005. 11

[15] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000. 9

[16] T.-P. Tian, R. Li, and S. Sclaroff. Articulated pose estimation in a learned smooth space of feasible solutions. In *CVPR*, 2005. 1, 9, 11

[17] R. Urtasun, D. J. Fleet, A. Hertzmann, and P. Fu. Priors for people tracking from small training sets. In *ICCV*, Beijing, China, 2005. 1, 9, 12

[18] Q. Wang, G. Xu, and H. Ai. Learning object intrinsic structure for robust visual tracking. In *CVPR (2)*, pages 227–233, 2003. 9

[19] C. K. I. Williams and D. Barber. Bayesian classication with Gaussian processes. *PAMI*, 20(12):1342–1351, 1998. 4, 8, 9