Eurographics/ ACM SIGGRAPH Symposium on Computer Animation (2010)
M. Otaduy and Z. Popovic (Editors)

BoLeRO: A Principled Technique for Including Bone Length Constraints in Motion Capture Occlusion Filling

Lei Li, James McCann, Nancy Pollard and Christos Faloutsos

Carnegie Mellon University

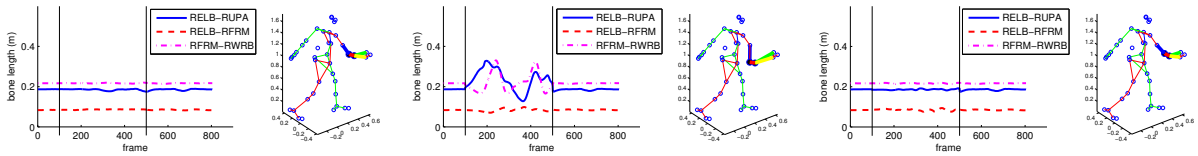


Figure 1: *Reconstructing two marker positions on the right arm from frame 100 to 500 of a walking motion (#132.43). Graphs show bone lengths over time; markers are stills at frame 241. LDS/DynaMMo (textbfmiddle) fails to preserve inter-marker distances present in the original motion (**left**). We propose BoLeRO which does much better (**right**).*

Abstract

Given a motion capture sequence with occlusions, how can we recover the missing values, respecting bone-length constraints? Recent past work uses Linear Dynamical Systems (LDS), which work well, except for occasionally violating such constraints, and thus lead to unrealistic results. Our main contribution is a principled approach for preserving such distances. Specifically (a) we show how to formulate the problem as a constrained optimization problem, using two variations: hard constraints, and soft constraints; (b) we show how to efficiently solve both variations; (c) we demonstrate the realism of our approaches against competitors, on real motion capture data, illustrating that our 'soft constraints' version eventually produces more realistic results.

1. Introduction

Given motion capture data, with occlusion, how can we recover the missing values, so that we obey bone-length constraints?

Optical motion capture is a useful method for computer animation. In this technique, cameras are used to track reflective markers on an actor's body, and the pose of the actor is reconstructed from these marker positions. Of course, such systems are not infallible, and inevitably some markers cannot be tracked due to occlusions or awkward camera placement. Similarly, one could imagine concatenating two such sequences of marker motion and treating the "transition" region as a large tracking failure.

Table 1: *Comparison of Occlusion Filling Methods*

Advantages Method	Bone Length constraints	Black-out
Spline	×	✓
MSVD	×	×
LDS/DynaMMo	×	✓
BoLeRO	✓	✓

Currently such occlusions are filled manually or through ad-hoc methods. Straightforward methods, like linear interpolation and spline interpolation, do not agree with human intuition, giving poor results. A more principled approach to occlusion filling would be to use a statistical model that accounts for correlations between markers and dynamics across time. One intuitive formulation is a linear dynamical system (LDS), which models observed data as noisy linear projections of low-dimensional state which evolves via noisy linear dynamics.

One problem with using LDS in this setting is that they do not preserve inter-marker distances. While a joint-angle rep-

resentation would solve this problem, it would both require that a skeleton be fit to the data (which would prevent LDS from being used for occlusion filling), and it would present a weight-selection challenge (a small angle error in the shoulder is much more noticeable than a small angle error in the wrist).

We show how to solve this problem in an explicit and principled manner, by specifying inter-marker distance constraints and learning an LDS that operates in this constrained space. The focus of our work is to handle occlusions automatically, agreeing with human intuition. Ideally we would like a method with the following properties:

1. Bone Length Constraint: It should be able to keep relative distance for markers on the same bone.
2. Black-out: It should be able to handle "black-outs", when all markers disappear (e.g., a person running behind a wall, for a moment).

Additionally, we also want our method to be scalable and automatic, requiring few (and, ideally, zero) parameters to be set by a human.

In this paper, we propose BoLeRO (**Bo**ne **le**ngth constrained **r**econstruction for **o**cclusion). Fig. 2 shows the reconstructed signal for an occluded running motion. Our method gives the best result close to the original value. Our main idea is to simultaneously exploit two properties: (a) body rigidness, through the bone length constraints; and (b) motion smoothness, by using the dynamical system to keep the moving trend. This two-prong approach can help us handle even "*black-outs*", which we define as time intervals where we lose track of all the markers.

The main contributions of this paper are as follows:

1. We setup the occlusion filling problem and formulate the bone length constraints in a principled way.
2. We propose effective algorithms (Expectation-Maximization-Newton/Gradient) to solve the problem, yielding results agreeing with human intuition.
3. We perform experiments on real motion capture sequences to demonstrate the additional benefits from enforcing the bone length constraints.

The rest of the paper is organized as follows: In Section 2, we review the related work; the proposed method and its discussion are presented in Section 3; the experimental results are presented in Section 4. We conclude the paper in Section 5.

2. Related work

Past occlusion filling methods and related techniques for occlusion filling can be classified into the following categories: (1) interpolation methods; (2) skeleton based [HFP*00, ZVDH03]; (3) dimensionality reduction and latent variables [LM06, PH06, THR07]; (4) database backed [HGP04, CH05]; (5) dynamical systems [WFH08, LMPF09].

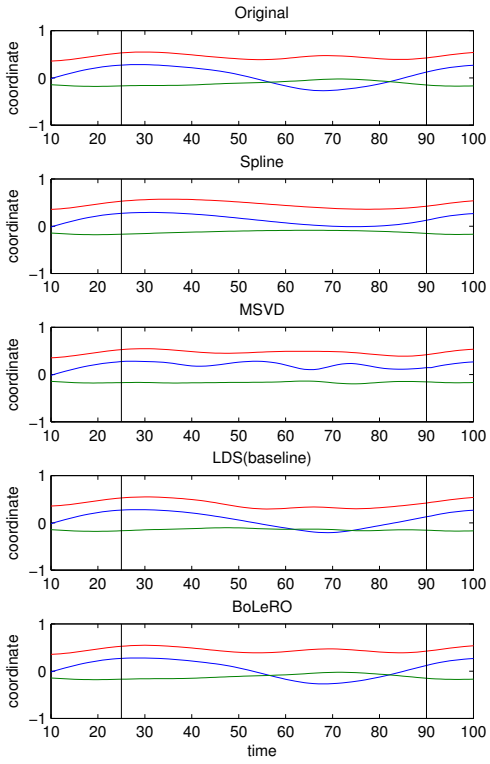


Figure 2: *Original and reconstructed xyz-coordinates of the marker on right knee for a running motion (subject#127.07) with occlusion from time 25 to 90 (marked with vertical lines). This is 45% of 145 frames in total. x,y,z are in blue, green and red respectively. Top to bottom figures correspond to original motion, reconstruction from spline, MSVD, LDS and our proposed BoLeRO, respectively.*

Interpolation methods: Linear interpolation and splines, are commonly used in time series smoothing and also motion capture systems to handle missing markers. These interpolation methods are generally effective for short period occlusions or occasional missing markers.

Skeleton based methods: Herda et al [HFP*00] used a human body skeleton to track and reconstruct the 3-d marker positions. Their method could predict the position using three previous markers by calculating the kinetics, when a marker is missing. Markers in motion capture system are usually captured in 3D space, while many applications work in joint angle space, thus a mapping from raw 3D data to joint angles is often required. Instead of fixed skeleton, Kirk et al [KOF05] proposed a method to automatically construct structural skeleton from motion capture data. Zordan and

Van Der Horst [ZVDH03] used a fixed limb-length skeleton to map the markers on full body to a representation in joint angles plus reference body center. Our proposed approach works directly in 3D space, there it does not require such mapping. These skeleton methods could work well for short segment of occlusions, however our method could handle much longer occlusions, as well as black-outs.

Dimensionality reduction and latent variable models: Liu and McMillan [LM06] proposed a method that projects the markers into linear principal components and reconstructs the missing parts from the linear models. This approach is similar to the Missing Value SVD (MSVD) [NJ03] with single iteration. Furthermore, they proposed an enhanced Local Linear method from a mixture of such linear models. Park and Hodgins [PH06] also used PCA to estimate the missing markers for skin deformation capturing. There were also work on nonlinear models for human motion. Taylor et al [THR07] used a conditional restricted Boltzmann machine (CRBM) with discrete hidden states to model human motion. Their approach could learn non-linear binary representations for a motion frame, conditioned on several previous frames. Therefore, their method could fill in the missing values from the prediction of several previous frames.

Database backed approach: Hsu et al [HGP04] proposed a method to map from a motion control specification to a target motion by searching over patterns in existing database. Chai and Hodgins [CH05] used a small set of markers as control signals and reconstruct the full body motion from a pre-recorded database. These methods could also generate motions follow the bone lengths, however they are repetitions or interpolated motions taken directly from database, thus could not generate new motions. The subset of markers should be known in advance, while our method does not assume fixed subsets of the dimension observed or missing.

Dynamical systems: Previously, Kalman filters were used for tracking system [DU03] with carefully defined parameters. Shumway and Stoffer [SS82] proposed EM algorithm to learn the model parameters. Wang et al [WFH08] took a nonparametric approach to model human motion and proposed a Gaussian process dynamical model, which includes a chain of latent variables and nonlinear mapping from the latent space to observed motion. In case of missing observation, it could use the learned model to estimate expectation of missing markers. Aristidou et al [ACL08] used Kalman filters to predict the missing markers, with parameters conforming to Newton dynamics. Recently, Li et al [LMPF09] used Linear Dynamical Systems to model motion capture data, and proposed an algorithm to recover the missing values. We will use it as the baseline method for comparison and describe it in more details in Section 3.1.

Liu and McMillan [LM06] provide a nice summary of related work on occlusion for motion capture data as well as of techniques for related tasks such as motion tracking.

Comparing against all these methods, our proposed BoLeRO can (a) capture the coupling between multiple markers like dimensionality reduction does; (b) it can generate motions that follow the dynamics of natural human motion like LDS/Kalman filters do; (c) it can enforce inter marker distances for those on the same bone (exactly or with a small toleration) as the skeleton method does; (d) it models the motion as a whole, instead of treating each frames individually, and thus BoLeRO is able to use the observed portion as much as possible. That is, each previous method exhibits only one or two of the above properties, but not all of them. This is the intuition why our BoLeRO method achieves better recovery of occlusions.

3. Proposed Method: BoLeRO

A typical motion capture system usually uses optical cameras to track passive markers in 3-D space. Mathematically, the observations of marker positions form a multi-dimensional time series, denoted as $\mathcal{X}$ (a T $\times m$ matrix). In case of marker occlusion (denoted as $\mathcal{X}_m$, the set of variables that are missing from the T $\times m$ matrix $\mathcal{X}$), the goal is to fill in the blanks to reconstruct most natural motion according to the human eye. The relevant symbols are described in Table 2.

Our method is motivated by observing the important role of rigid bones in human body that preserve relative positions of markers: markers on the same bone will maintain a given distance between them. We capture these fixed distances through bone length constraints. Our approach addresses the occlusion filling problem by enforcing the bone length constraints on top of a traditional dynamical system for motion capture data.

How should we incorporate BLC into the dynamic systems? There are two choices with varying implications: namely, hard constraints and soft constraints. The former exerts exact inter-marker distances for same bone markers, while the latter general follows the constraint while allowing occasional violations. We will describe each of them and compare their effectiveness in later sections. Here we first define the meaning of bone length in our context.

Definition 1 A set B lists bone length constraints (BLC), and it contains the following elements:

$$B = \{\langle i, j, d_{i,j}\rangle | \text{marker } i, j \text{ on the same bone}\}$$

where $d_{i,j}$ is the distance between marker i and j.

Definition 2 A matrix W is said to indicate the missing observation if

$$W(t,i) = \begin{cases} 1 & i\text{-th marker missing at time } t, \\ 0 & i\text{-th marker observed at time } t \end{cases}$$

Definition 3 A pair of marker i and j are said to conform to the BLC B if their coordinates of all time tick follows

$$\langle i, j, d_{i,j} \rangle \in B \Rightarrow \|\mathbf{x}_t^{(i)} - \mathbf{x}_t^{(j)}\|^2 = d_{i,j}^2$$

where $\mathbf{x}_t^{(i)}$ denotes the coordinates for i-th marker at time t.

Table 2: *Symbols, Acronyms and Definitions*

Symbol	Definition
$\mathcal{X}$	a motion sequence with missing values $(\mathbf{x}_1, ...\mathbf{x}_T)$ of T time-ticks in m dimensions
$\mathcal{X}_g$	the observed values in a motion sequence $\mathcal{X}$
$\mathcal{X}_m$	variables for the missing values in a motion sequence $\mathcal{X}$
$\mathbf{x}_t^{(i)}$	position of marker i at time t
$\mathbf{z}_t$	hidden variables at time t
W	01 matrix indicating missing values (1=missing)
m	number of dimensions (e.g., marker positions) - m=123 here
H	number of hidden dimensions
BLC	bone length constraint
EMN/EMG	expectation, maximization and Newton/gradient descent

3.1. Background

Linear Dynamical Systems (LDS, also known as Kalman filters) are commonly used in motion tracking systems. The basic idea is to identify the hidden variables (e.g. velocity, acceleration) in observed data (marker positions) and build a statistical model to characterize the transitions of hidden variables. Such models could then reproduce the motion dynamics, as well as the correlations among markers by choosing a proper number of hidden variables. In LDS, a multi-dimensional motion sequence is modeled with a hidden Markov chain, as follows.

$$\mathbf{z}_1 = \mu_0 + \omega_0 \quad (1)$$

$$\mathbf{z}_{n+1} = \mathbf{F} \cdot \mathbf{z}_n + \omega_n \quad (2)$$

$$\mathbf{x}_n = \mathbf{G} \cdot \mathbf{z}_n + \varepsilon_n \quad (3)$$

where $\Theta = \{\mu_0, \Gamma, \mathbf{F}, \Lambda, \mathbf{G}, \Sigma\}$ is the set of parameters. μ_0 is initial state of the whole system, $\mathbf{x}_n$ and $\mathbf{z}_n$ denote marker coordinates and hidden variables at time n, respectively. $\mathbf{F}$ implies the transition dynamics and $\mathbf{G}$ is the observation projection. ω_0, ω_i and $\varepsilon_i (i = 1 \ldots T)$ are multivariate Gaussian noises with the following distributions:

$$\omega_0 \sim \mathcal{N}(0, \Gamma) \;\; \omega_i \sim \mathcal{N}(0, \Lambda) \;\; \varepsilon_j \sim \mathcal{N}(0, \Sigma) \quad (4)$$

In the case of missing observations, Li et al [LMPF09] proposed DynaMMo, an expectation-maximization (EM) like algorithm to recover the marker positions by estimating the expectation of occluded values, given the observed parts, $\mathbb{E}[\mathcal{X}_m | \mathcal{X}_g]$. Their algorithm finds solutions to maximize the expected log-likelihood with respect to the model parameters, the hidden variables and the missing observations as well. However, their method would have a hard time for multiple markers of long occlusions as we point out the experiments.

3.2. BoLeRO-HC (hard constraints)

Intuition: The traditional method to estimate the missing values in the sequence data is to minimize a squared loss function on Eq (1)-(4), penalizing the model complexity. While a LDS based method such as DynaMMo [LMPF09] could recover short occlusion, it suffers in cases of longer and multiple marker occlusions, because reconstructed markers may break rigid bodies and violate bone length constraints. Our proposed method is based on the basic intuition on human motion: those markers attached to the same bones should not fall apart. Our main contribution in the current paper is to demonstrate the usefulness of domain knowledge, in this case bone length constraints in occlusion filling. Following this intuition, we propose BoLeRO-HC (hard constraints), enforcing the exact bone length constraints in a LDS-based model. We make a first conjecture, bone length constraints will help recover motion occlusions better, in addition to traditional dynamics information as modeled by LDS. Here we first formulate the domain knowledge and define the bone length constraints.

Problem formulation: We will first present the proposed BoLeRO-hard constraint here. With the bone length constraints (BLC), we link the naturalness of a motion to whether it conforms to desired bone lengths under temporal movement. In our model, we assume the motion is moving according to the linear dynamical systems. Given an occluded motion sequence $\mathcal{X}$, occlusion indication matrix W and an additional BLC B, the problem is to fill in the occlusion so that the resulting motion follows the moving dynamics as captured by LDS and conforms to the bone length constraints. Our proposed BoLeRO will recover the missing markers by estimating the "good" expectation $\mathbb{E}[\mathcal{X}_m | \mathcal{X}_g]$, which conforms to the inter-marker distances on the same bone. The sequence of marker coordinates are modeled based on LDS as above (Eq. (1)-(3)) with additional constraints. Mathematically, the occlusion filling problem and its cost function are defined as follows.

Problem 1 (BoLeRO, hard constraints) Given (a) $\mathcal{X}_g$ (the observed marker positions), (b) B (bone length constraint), and (c) occlusion indication matrix W, find Θ and X_m to solve the following objective function:

$$\min \quad Q(X_m, \Theta) \quad (5)$$

$$\text{subject to} \|x_t^{(i)} - x_t^{(j)}\|^2 - d_{i,j}^2 = 0 \quad \forall \langle i, j, d_{i,j} \rangle \in B$$

with the objective function $Q(\cdot)$ to be

$$\begin{aligned} Q(X_m,\Theta) = &\frac{1}{2}\mathbb{E}[(\mathbf{z}_1-\mu_0)^T\Gamma^{-1}(\mathbf{z}_1-\mu_0) \\ &+\sum_{t=2}^{T}(\mathbf{z}_t-\mathbf{F}\cdot\mathbf{z}_{t-1})^T\Lambda^{-1}(\mathbf{z}_t-\mathbf{F}\cdot\mathbf{z}_{t-1}) \\ &+\sum_{t=1}^{T}(\mathbf{x}_t-\mathbf{G}\cdot\mathbf{z}_t)^T\Sigma^{-1}(\mathbf{x}_t-\mathbf{G}\cdot\mathbf{z}_t)] \\ &+\frac{\log|\Gamma|}{2}+\frac{T-1}{2}\log|\Lambda|+\frac{T}{2}\log|\Sigma| \end{aligned}$$

where $\mathbf{x}_t^{(i)}$ is coordinates of i-th marker at time t, and Θ denotes model parameters $\Theta=\{\mathbf{F},\mathbf{G},\mathbf{z}_0,\Gamma,\Lambda,\Sigma\}$.

Algorithm: The main goal of the learning algorithm is to find optimal values for $\mathcal{X}_m$ such that the objective function $Q(\mathcal{X}_m,\Theta)$ is minimized under the bone length constraints. To solve the optimization problem, we observe that the constraints have nothing to do with the hidden variables and model parameters Θ, which suggests the following coordinate-descent style algorithm. At the high level, our proposed algorithm optimizes the parameters and unknowns piece-wisely and iteratively through the "EMN" procedure (*E*xpectation, Maximization, and Newton optimization), as shown in Algorithm 1: We use Expectation-Maximization [LMPF09] to estimate the posterior distribution $P(\mathcal{Z}|\mathcal{X})$, its sufficient statistics ($\mathbb{E}[\mathbf{z}_t]$, $\mathbb{E}[\mathbf{z}_t\mathbf{z}_t^T]$ and $\mathbb{E}[\mathbf{z}_t\mathbf{z}_{t-1}^T]$), and Θ respectively (steps 1 and 2 in Algorithm 1), and fill in missing values using Newton's method to solve the Lagrange derived from the BLC (step 3 in Algorithm 1). Finally, we update the model parameters Θ by maximizing the log likelihood (i.e. minimizing Q), and iterate until convergence.

In more detail, our proposed BoLeRO-HC uses Lagrange multipliers to handle constraints for frames with missing values. The Lagrangian is given by:

$$\begin{aligned} \mathcal{L}(\mathcal{X}_m,\Theta,\eta) = &\frac{1}{2}\mathbb{E}[(\mathbf{z}_1-\mathbf{z}_0)^T\Gamma^{-1}(\mathbf{z}_1-\mathbf{z}_0) \\ &+\sum_{t=2}^{T}(\mathbf{z}_t-\mathbf{F}\mathbf{z}_{t-1})^T\Lambda^{-1}(\mathbf{z}_t-\mathbf{F}\mathbf{z}_{t-1}) \\ &+\sum_{t=1}^{T}(\mathbf{x}_t-\mathbf{G}\mathbf{z}_t)^T\Sigma^{-1}(\mathbf{x}_t-\mathbf{G}\mathbf{z}_t) \\ &+\frac{1}{2}\log|\Gamma|+\frac{T-1}{2}\log|\Lambda|+\frac{T}{2}\log|\Sigma|] \\ &+\sum_{t=1}^{T}\sum_{\langle i,j,d_{i,j}\rangle\in B}\eta_{tij}(\|\mathbf{x}_t^{(i)}-\mathbf{x}_t^{(j)}\|^2-d_{i,j}^2) \end{aligned} \tag{6}$$

where $\eta=\{\eta_{tij}\}$ are Lagrange multipliers. Note here we also include the dummy Lagrange multipliers for observed markers, however, because since those marker positions are known, it will not affect the result.

To derive a solution for the constrained optimization problem, we follow the "EMN" guideline, expectation ($P(\mathcal{Z}|\mathcal{X})$), maximization ($\Theta$), and Newton optimization ($\mathcal{X}_m$): we first take derivative of $\mathcal{L}$ with respect to Θ, yielding the forward-backward belief propagation (also known as Kalman filtering and smoothing) for expectation step and maximization equations. Since Θ is not involved in constraints, the resulting updating equations are the same as those in [LMPF09] and derived in Appendix A.

To estimate $\mathcal{X}_m$ and η's, we use Newton's method to iteratively search and minimize the objective function with respect to the constraints. The optimal solution is specified by critical point, which requires $\frac{\partial\mathcal{L}}{\partial\mathcal{X}_m}=0$ and $\frac{\partial\mathcal{L}}{\partial\eta}=0$. In this step, we iteratively update the $\mathcal{X}_m$ and η in the Newton's descent direction. Let $\mathbf{x}_t^M$ and η_t denote the unobserved marker positions and Lagrangian multipliers at time t, respectively. During each iteration, we update them according to the following update rules:

$$\begin{pmatrix}\mathbf{x}_t^M\\ \eta_t\end{pmatrix}^{new} \longleftarrow \begin{pmatrix}\mathbf{x}_t^M\\ \eta_t\end{pmatrix} - \alpha(\nabla^2_{\mathbf{x}_t^M,\eta_t}L)^{-1}\cdot\nabla_{\mathbf{x}_t^M,\eta_t}\mathcal{L} \tag{7}$$

where $\nabla_{\mathbf{x}_t^M,\eta_t}\mathcal{L}$ is the partial gradient and $\nabla^2_{\mathbf{x}_t^M,\eta_t}\mathcal{L}$ is the Hessian matrix. Their detailed derivations are given in appendix B.

3.3. BoLeRO-SC (soft constraints)

Intuition - Motivation The hard constraint formulation above would be ideal, except that in reality, markers slightly move, and reality itself *violates* the bone length constraints (BLC)! In such cases, hard constraints may land to a solution with abrupt discontinuity in the recovered marker position, albeit the corresponding marker distances are well preserved.

This implies we do not have to underscore the exact preservation of the bone length, while the ideal system should allow some "reasonable" variation. The intuition comes naturally, the recovered missing values should be a trade off between approaching the maximum likelihood and preserving bone length. To alleviate this problem, we relax the bone constraints and instead solve the following soft constrained problem. Our objective is the likelihood, with additional penalty on the deviation from the desired bone length of those missing markers on the same bones.

Problem Formulation Following the intuition, we get the following objective function: the negative log likelihood pe-

Algorithm 1: Proposed EMN algorithm for BoLeRO-HC

Input: X_g, B (BLC), W (missing indication matrix), H (hidden dimension)
Output: Recovered motion sequence: $\hat{X}$

```
// initialization
```
$X_m \leftarrow 0$; `// or other initialization`
$\hat{X} \leftarrow X_g \cup X_m$;
$\mathbf{F}, \mathbf{G}, \mu_0 \leftarrow$ random values;
$\Gamma, \Lambda, \Sigma \leftarrow I$;
$\Theta \leftarrow \{\mathbf{F}, \mathbf{G}, \mu_0, \Gamma, \Lambda, \Sigma\}$;
repeat
1 E-step: estimate the posterior $P(Z|\hat{X};\Theta)$ and its sufficient statistics $\mathbb{E}[\mathbf{z}_t|\hat{X};\theta]$, $\mathbb{E}[\mathbf{z}_t{\mathbf{z}_t}^T|\hat{X};\theta]$, and $\mathbb{E}[\mathbf{z}_t{\mathbf{z}_{t-1}}^T|\hat{X};\theta]$ using belief propagation (Eq.(15-30));
2 M-step: Minimizing Eq (6) with respect to Θ (Eq.(32-36))

$$\Theta^{new} \leftarrow \arg\min_{\Theta} \mathcal{L}(X_m, \Theta, \eta)$$

3 **for** $t \leftarrow 1$ *to* T **do**
`// N-step: estimating missing using Newton's method`
$k \leftarrow$ number of missing markers at time t;
$\eta_t \leftarrow (\underbrace{0, \ldots, 0}_{k})^T$;
$\alpha \leftarrow 1/2$; `// step size`
repeat
$D \leftarrow \nabla_{\mathbf{x}_t^M, \eta_t} \mathcal{L}(\cdot)$; `// Page.10 Eq.(37)`
$H \leftarrow \nabla^2_{\mathbf{x}_t^M, \eta_t} \mathcal{L}(\cdot)$; `// Page.10 Eq.(38)`
$\mathbf{y} \leftarrow \hat{X}_t^{(i)} \quad \forall i.W_{t,i} = 1$;
$\begin{pmatrix} \mathbf{y} \\ \eta_t \end{pmatrix}^{new} \longleftarrow \begin{pmatrix} \mathbf{y} \\ \eta_t \end{pmatrix} - \alpha H \cdot D$
$\hat{X}_t^{(i)} \leftarrow \mathbf{y} \quad \forall i.W_{t,i} = 1$;
until *converge* ;
until *converge* ;

nalized by the deviation from the bone length constraints.

$$\begin{aligned}
\min \quad & f(X_m, \Theta) \qquad (8)\\
=& \frac{1}{2}\mathbb{E}\Big[(\mathbf{z}_1 - \mu_0)^T \Gamma^{-1} (\mathbf{z}_1 - \mu_0)\\
&+ \sum_{t=2}^{T} (\mathbf{z}_t - \mathbf{F}\cdot\mathbf{z}_{t-1})^T \Lambda^{-1} (\mathbf{z}_t - \mathbf{F}\cdot\mathbf{z}_{t-1})\\
&+ \sum_{t=1}^{T} (\mathbf{x}_t - \mathbf{G}\cdot\mathbf{z}_t)^T \Sigma^{-1} (\mathbf{x}_t - \mathbf{G}\cdot\mathbf{z}_t)\Big]\\
&+ \frac{\log|\Gamma|}{2} + \frac{T-1}{2}\log|\Lambda| + \frac{T}{2}\log|\Sigma|\\
&+ \frac{\lambda}{2}\sum_{t=1}^{T}\sum_{\langle i,j,d_{i,j}\rangle \in B} (W_{t,i}|W_{t,j})(\|\mathbf{x}_t^{(i)} - \mathbf{x}_t^{(j)}\|^2 - d_{i,j}^2)^2
\end{aligned}$$

where $W_{t,i}|W_{t,j} = W_{t,i} + W_{t,j} - W_{t,i}W_{t,j}$.

Algorithm: To solve the optimization problem, we propose a coordinate descent approach, alternatingly optimizing over a set of unknown variables or parameters (see Algorithm 2):

1. E-step: fix Θ and missing $\mathcal{X}_m$, using Kalman filtering and Kalman smoothing to estimate posterior $P(\mathcal{Z}|\mathcal{X};\Theta)$,
2. M-step: update the model parameters Θ,
3. G-step: Fix Θ, estimate the missing $\mathcal{X}_m$ under soft constraints using gradient descent, with the previously computed $P(\mathcal{Z}|\mathcal{X};\Theta)$.

By taking partial derivatives over Θ and setting to zero, we obtain the same equations for E-step and M-step as in above EMN for hard constraints. While N-step is replaced with the gradient descent on soft constraints. The update rule for G-step is:

$$\mathbf{x}_t^{(i)} \leftarrow \mathbf{x}_t^{(i)} - \alpha \cdot \frac{\partial f}{\partial \mathbf{x}_t^{(i)}} \qquad (9)$$

where $x_t^{(i)}$ denotes the i-th marker coordinates at time t.

$$\begin{aligned}
&\frac{\partial f}{\partial \mathbf{x}_t^{(i)}} = I_{(i)} \cdot \Sigma^{-1} \cdot (\mathbf{x}_t - \mathbf{G}\cdot\mathbb{E}[\mathbf{z}_t])\\
&+2\lambda \sum_{\langle i,j,d_{i,j}\rangle \in B} (W_{t,i}|W_{t,j})(\|\mathbf{x}_t^{(i)} - \mathbf{x}_t^{(j)}\|^2 - d_{i,j}^2)(\mathbf{x}_t^{(i)} - \mathbf{x}_t^{(j)})
\end{aligned} \qquad (10)$$

The learning rate depends on the proper choosing of the learning step size α. We developed an adaptive scheme for adjusting α according to value of the objective function. The basic idea is to enlarge α whenever the objective decreases and to shrink α whenever it increases.

To make the scheme work, we observe that the partial derivative $\frac{\partial f}{\partial \mathbf{x}_t^{(i)}}$ is independent of all the rest time ticks. So in our algorithm, we isolate the optimization for each time tick, and adaptively choose the learning rate for that time tick. Specifically, we define the following time-decomposed objective function:

$$\begin{aligned}
f_t(x_t) =& \frac{1}{2}\mathbb{E}[(\mathbf{x}_t - \mathbf{G}\cdot\mathbf{z}_t)^T \Sigma^{-1} (\mathbf{x}_t - \mathbf{G}\cdot\mathbf{z}_t)]\\
&+ \frac{\lambda}{2}\sum_{\langle i,j,d_{i,j}\rangle \in B} (W_{t,i}|W_{t,j})(\|\mathbf{x}_t^{(i)} - \mathbf{x}_t^{(j)}\|^2 - d_{i,j}^2)^2
\end{aligned} \qquad (11)$$

Observing $\frac{\partial f_t}{\partial \mathbf{x}_t^{(i)}} = \frac{\partial L}{\partial \mathbf{x}_t^{(i)}}$, the update rule becomes

$$\mathbf{x}_t^{(i)} \leftarrow \mathbf{x}_t^{(i)} - \alpha \cdot \frac{\partial f_t}{\partial \mathbf{x}_t^{(i)}} \qquad (12)$$

The adaptive gradient descent method works as follows: it only accepts the update when the update will decrease the time-decomposed objective function f_t, doubling α in this case; otherwise halving α.

Algorithm 2: Proposed EMG algorithm for BoLeRO-SC

Input: X_g, B (BLC), W (missing indication matrix), H (hidden dimension)
Output: Recovered motion sequence: $\hat{X}$

```
// initialization
```
$X_m \leftarrow 0$; `// or other initialization`
$\hat{X} \leftarrow X_g \cup X_m$;
$\mathbf{F}, \mathbf{G}, \mu_0 \leftarrow$ random values;
$\Gamma, \Lambda, \Sigma \leftarrow I$;
$\Theta \leftarrow \{\mathbf{F}, \mathbf{G}, \mu_0, \Gamma, \Lambda, \Sigma\}$;
repeat
1 E-step: estimate the posterior $P(Z|\hat{X};\Theta)$ and its sufficient statistics $\mathbb{E}[\mathbf{z}_t|\hat{X};\theta]$, $\mathbb{E}[\mathbf{z}_t {\mathbf{z}_t}^T|\hat{X};\theta]$, and $\mathbb{E}[\mathbf{z}_t {\mathbf{z}_{t-1}}^T|\hat{X};\theta]$ using belief propagation (Eq.(15-30));
2 M-step: Minimizing Eq (8) with respect to Θ (Eq.(32-36))

$$\Theta^{new} \leftarrow \arg\min_{\Theta} f(X_m, \Theta)$$

3 **for** $t \leftarrow 1$ *to* T **do**
`// G-step: estimating missing using gradient descent`
$\alpha \leftarrow 1$; `// step size`
repeat
foreach *i s.t.* $W(t,i) = 1$ **do**
$D \leftarrow \frac{\partial f}{\partial \mathbf{x}_t^{(i)}}$; `// Eq.(10)`
$(\mathbf{x}_t^{(i)})^{new} \leftarrow \mathbf{x}_t^{(i)} - \alpha \cdot D$
if $f_t((\mathbf{x}_t)^{new}) < f_t(\mathbf{x}_t)$ **then** accept update
update $\hat{X}$ with $(\mathbf{x}_t)^{new}$;
$\alpha \leftarrow 2\alpha$;
elsereject update
$\alpha \leftarrow \frac{\alpha}{2}$;
until *converge* ;
until *converge* ;

4. Experimental Results

We performed experiments on real human motion capture data to evaluate the effectiveness of our proposed method.

We used a public dataset from CMU mocap database [CMU]. Each motion consists of 200 to 1500 frames and 123 features of marker positions (41 markers), converted to body local coordinates by estimating root position and body facing. We rescaled units into meters, which will improve the computation stability since all numbers are within the range of [-2, 2].

For each of the motion in our trial, we create its bone length constraints, B, by estimating the average inter-marker distances (e.g. marker RTHI and RSHN are on the same bone). Alternatively, we can construct the BLC by estimating the variance of the inter marker distances and thresholding, or algorithms by Kirk et al [KOF05] and de Aguiar [dATS06], however bone length estimation is beyond the scope of our paper. For both baseline and BoLeRO, we set the hidden dimension $H = 16$, which is corresponding to over 95% of energy in original data. We set $\lambda = 10^6$ for BoLeRO-SC in our experiments.

To evaluate the effectiveness of our proposed methods BoLeRO-HC and BoLeRO-SC, we select a trial set with 9 motions representing a variety of motion types, including running, walking, jumping, sports, and martial art. We did a statistical test as well as case studies. In statistical test, we randomly occluded a marker for a random consecutive segment, and tested the reconstruction with all candidate methods. Each trial is repeated 10 times with a different random occlusion. Fig. 4 shows reconstruction mean square error (Eq. 13) against the original motion. Notice BoLeRO-SC consistently has lower mse over the baseline LDS/DynaMMo while BoLeRO-HC occasionally does.

$$\text{mse} = \frac{\sum_{t,i} W_{t,i}(\hat{X}_{t,i} - X_{t,i}^{true})^2}{\sum_{t,i} W_{t,i}} \tag{13}$$

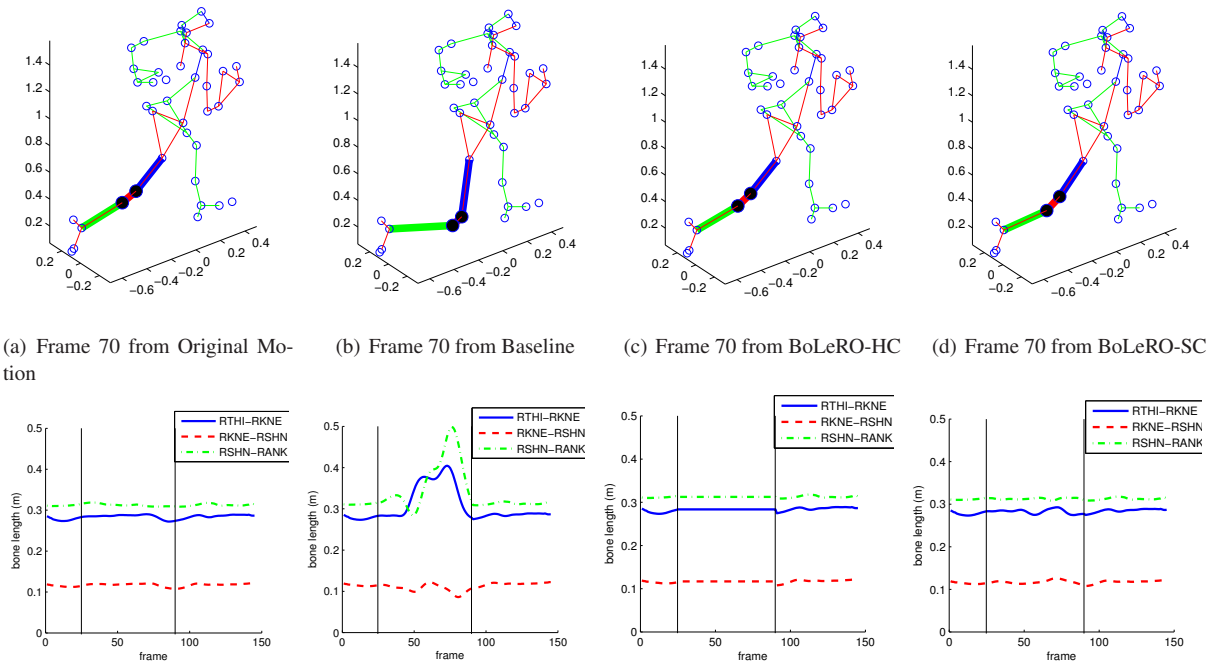
where W indicates missing markers (if $= 1$).

We also test the methods on two or more markers missing as well. Fig. 3 shows a case of running motion (subject#127.07) and reconstruction results by two methods. Two markers on the right leg (RKNE for knee and RSHN for shin) are occluded from frame 25 to 90 inclusive. Fig. 2 shows the time plot of coordinates for one marker on the right knee, where spline and MSVD clearly deviates much from the original data, hence we did not include them in the following distance plot. Fig. 3(e)-3(g) show the distances between the two markers and adjacent markers on the body skeleton (thigh-to-knee, knee-to-shin and shin-to-ankle in blue, red and green respectively). All three distances should ideally be constant. The result generated by baseline method violates the bone constraint (particularly around frame 70), while BoLeRO clearly improves the quality of reconstruction by obeying the corresponding BLC. Additional results and animations are shown in the accompanying video.

5. Conclusions

We focus on the problem of occlusion in motion capture data, and specifically on the reconstruction so that we obey bone length constraints.

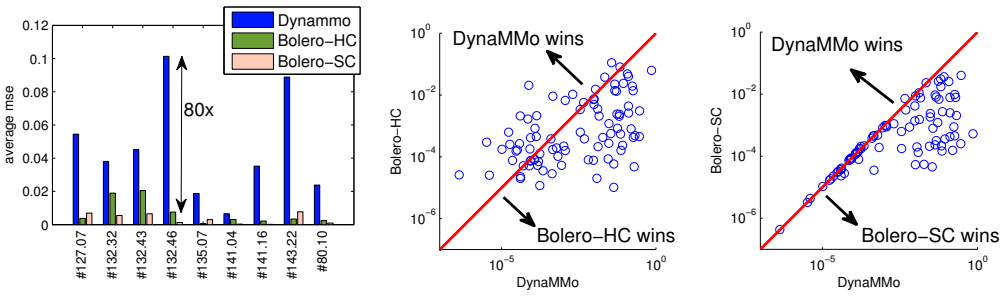
Motion capture is a useful technique to obtain realistic motion animation, where optical cameras are used to track marker movement on actors. Unfortunately markers will be out of view sometimes, especially in full body motions like running, football playing and bounce walking, etc., and it takes hours/days for human experts to manually fix the gaps. How can we handle the occluded motion and fill in the gaps automatically and effectively, while respecting bone-length



(a) Frame 70 from Original Motion (b) Frame 70 from Baseline (c) Frame 70 from BoLeRO-HC (d) Frame 70 from BoLeRO-SC

(e) Original bone length (f) Bone length from baseline (g) Bone length from BoLeRO-HC (h) Bone length from BoLeRO-SC

Figure 3: *Recovery result for an occluded running motion (subject #127.07). Two markers on the right leg (RKNE and RSHN, marked with solid circles in top) are occluded from frame 25 to 90 (marked with vertical lines in bottom). Top row: markers articulated with circles. Bold lines illustrate the bones of interest. Bottom row: bone lengths in original, baseline, BoLeRO-HC, and BoLeRO-SC, for thigh-to-knee (blue), knee-to-shin (red) and shin-to-ankle(green) respectively. Notice that the original, BoLeRO-HC, BoLeRO-SC lengths are near constant while the baseline has a severe violation of BLC.*



(a) average mse (b) DynaMMo v.s. BoLeRO-HC (c) DynaMMo v.s. BoLeRO-SC

Figure 4: *Comparison between baseline(LDS/DynaMMo), BoLeRO-HC, and BoLeRO-SC. 4(a) average mse for LDS/DynaMMo(blue), BoLeRO-HC(green), BoLeRO-SC(orange). 4(b),4(c) scatter plots of mse's in 90 trials on 9 motions. Our BoLeRO-SC consistently wins over DynaMMo (see (c) - all points are at or below diagonal), with a maximum of 80x improvement (see (a)), while BoLeRO-HC loses occasionally (see (b), points above diagonal).*

constraints? In this paper, we propose BoLeRO, a principled approach to reconstruct occluded motion using bone length constraints on body dynamics. The novelty is that it sets up the problem as a linear dynamical system with constraints, thus explicitly exploiting both the smoothness in motion dynamics, as well as the rigidness in distances between relevant markers. We give two versions of it: "hard constraints", and "soft constraints", where the reconstructed bone-lengths may slightly differ from the ideal ones.

The second contribution is that we propose a novel, fast algorithm to solve both versions of the problem, using our "EMN/EMG" formulation (expectation, maximization, Newton/gradient descent): The idea is to alternatingly estimate (a) the hidden variables (b) the parameters of the Linear Dynamical System and (c) the Lagrange multipliers (only in BoLeRO-HC) and missing values; and iterate until convergence.

Experiments on real data show that either version of BoLeRO are significantly better than straightforward alternatives (splines, linear interpolation) and they matches or outperforms very sophisticated alternatives like Kalman filters and recent missing-value algorithms [NJ03, LMPF09]. Among our two version, we recommend the "soft constraints" BoLeRO, which overwhelmingly outperforms the 'hard constraints' version.

References

[ACL08] ARISTIDOU A., CAMERON J., LASENBY J.: Predicting missing markers to drive real-time centre of rotation estimation. In *AMDO '08: Proceedings of the 5th international conference on Articulated Motion and Deformable Objects* (Berlin, Heidelberg, 2008), Springer-Verlag, pp. 238–247.

[CH05] CHAI J., HODGINS J. K.: Performance animation from low-dimensional control signals. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers* (New York, NY, USA, 2005), ACM, pp. 686–696.

[CMU] CMU: http://mocap.cs.cmu.edu.

[dATS06] DE AGUIAR E., THEOBALT C., SEIDEL H.: Automatic learning of articulated skeletons from 3d marker trajectories. pp. I: 485–494.

[DU03] DORFMÜLLER-ULHAAS K.: *Robust Optical User Motion Tracking Using a Kalman Filter*. Tech. Rep. 2003-6, Institut fuer Informatik, Universitätsstr. 2, 86159 Augsburg, May 2003.

[HFP*00] HERDA L., FUA P., PLÄNKERS R., D., BOULIC R., THALMANN D.: Skeleton-based motion capture for robust reconstruction of human motion. In *Computer Animation* (Philadelphia, PA, May 2000).

[HGP04] HSU E., GENTRY S., POPOVIĆ J.: Example-based control of human motion. In *SCA '04: Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation* (Aire-la-Ville, Switzerland, Switzerland, 2004), Eurographics Association, pp. 69–77.

[KOF05] KIRK A. G., O'BRIEN J. F., FORSYTH D. A.: Skeletal parameter estimation from optical motion capture data. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) 2005* (June 2005), pp. 782–788.

[LM06] LIU G., MCMILLAN L.: Estimation of missing markers in human motion capture. *Vis. Comput. 22*, 9 (2006), 721–728.

[LMPF09] LI L., MCCANN J., POLLARD N., FALOUTSOS C.: Dynammo: Mining and summarization of coevolving sequences with missing values. In *KDD '09: Proceeding of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA, 2009), ACM.

[NJ03] NATI N. S., JAAKKOLA T.: Weighted low-rank approximations. In *In 20th International Conference on Machine Learning* (2003), AAAI Press, pp. 720–727.

[PH06] PARK S. I., HODGINS J. K.: Capturing and animating skin deformation in human motion. *ACM Trans. Graph. 25*, 3 (2006), 881–889.

[SS82] SHUMWAY R. H., STOFFER D. S.: An approach to time series smoothing and forecasting using the em algorithm. *Journal of Time Series Analysis 3* (1982), 253–264.

[THR07] TAYLOR G. W., HINTON G. E., ROWEIS S. T.: Modeling human motion using binary latent variables. In *Advances in Neural Information Processing Systems 19*, Schölkopf B., Platt J., Hoffman T., (Eds.). MIT Press, Cambridge, MA, 2007, pp. 1345–1352.

[WFH08] WANG J. M., FLEET D. J., HERTZMANN A.: Gaussian process dynamical models for human motion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 30*, 2 (2008), 283–298.

[ZVDH03] ZORDAN V. B., VAN DER HORST N. C.: Mapping optical motion capture data to skeletal motion using a physical model. In *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation* (Aire-la-Ville, Switzerland, Switzerland, 2003), Eurographics Association, pp. 245–250.

Appendix A: Inference: Forward-backward estimation

Forward-backward Message Passing

Given the parameters $\Theta = (\mathbf{F}, \mathbf{G}, \mu_0, \Gamma, \Lambda, \Sigma)$, the estimation problem is finding the marginal distribution for hidden state variables given the observed data, e.g. $\hat{\mathbf{z}}_t = \mathbb{E}[\mathbf{z}_n \mid \mathcal{X}](n = 1, \ldots, \mathrm{T})$.

Assume the posterior up to current time tick is $p(\mathbf{z}_n|\mathbf{x}_1, \ldots, \mathbf{x}_n$, denoted by:

$$\hat{\alpha}(\mathbf{z}_n) = \mathcal{N}(\mu_n, \mathbf{V}_n) \tag{14}$$

We could obtain the following forward passing of the belief. The messages here are μ_n, $\mathbf{V}_n$ and $\mathbf{P}_{n-1}$(needed in later backward passing).

$$\begin{aligned}
\mathbf{P}_{n-1} &= \mathbf{F}\mathbf{V}_{n-1}\mathbf{F}^T + \Lambda && (15)\\
\mathbf{K}_n &= \mathbf{P}_{n-1}\mathbf{G}^T(\mathbf{G}\mathbf{P}_{n-1}\mathbf{G}^T + \Sigma)^{-1} && (16)\\
\mu_n &= \mathbf{F}\mu_{n-1} + \mathbf{K}_n(\mathbf{x}_n - \mathbf{G}\mathbf{F}\mu_{n-1}) && (17)\\
\mathbf{V}_n &= (\mathbf{I} - \mathbf{K}_n)\mathbf{P}_{n-1} && (18)\\
& && (19)
\end{aligned}$$

The initial messages are given by:

$$\begin{aligned}
\mathbf{K}_1 &= \Gamma\mathbf{G}^T(\mathbf{G}\Gamma\mathbf{G}^T + \Sigma)^{-1} && (20)\\
\mu_1 &= \mu_0 + \mathbf{K}_1(\mathbf{x}_1 - \mathbf{G}\mu_0) && (21)\\
\mathbf{V}_1 &= (\mathbf{I} - \mathbf{K}_1)\Gamma && (22)\\
& && (23)
\end{aligned}$$

For the backward passing, let $\gamma(\mathbf{z}_n)$ denote the marginal posterior probability $p(\mathbf{z}_n|\mathbf{x}_1, \ldots, \mathbf{x}_\mathrm{T})$ with the assumption:

$$\gamma(\mathbf{z}_n) = \mathcal{N}(\hat{\mu}_n, \hat{\mathbf{V}}_n) \tag{24}$$

The backward passing equations are:

$$\mathbf{J}_n = \mathbf{V}_n\mathbf{F}^T(\mathbf{P}_n)^{-1} \quad (25)$$

$$\hat{\mu}_n = \mu_n + \mathbf{J}_n(\hat{\mu}_{n+1} - \mathbf{F}\mu_n) \quad (26)$$

$$\hat{\mathbf{V}}_n = \mathbf{V}_n + \mathbf{J}_n(\hat{\mathbf{V}}_{n+1} - \mathbf{P}_n)\mathbf{J}_n^T \quad (27)$$

From the passed belief, we could obtain the following estimation:

$$\mathbb{E}[\mathbf{z}_n] = \hat{\mu}_n \quad (28)$$

$$\mathbb{E}[\mathbf{z}_n\mathbf{z}_{n-1}^T] = \mathbf{J}_{n-1}\hat{\mathbf{V}}_n + \hat{\mu}_n\hat{\mu}_{n-1}^T \quad (29)$$

$$\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^T] = \hat{\mathbf{V}}_n + \hat{\mu}_n\hat{\mu}_n^T \quad (30)$$

where the expectations are taken over the posterior marginal distribution $p(\mathbf{z}_n|\mathbf{x}_1,\ldots,\mathbf{x}_\mathrm{T})$.

Expectation-Maximization Learning

The new parameter Θ^{new} is obtain by maximizing $\mathcal{L}$ in Eq. 6 with respect to the components of Θ^{new} given the current estimate of Θ^{old}. Taking the derivatives and let them be zeros gives the following results:

$$\mu_0^{new} = \mathbb{E}[\mathbf{z}_1] \quad (31)$$

$$\Gamma^{new} = \mathbb{E}[\mathbf{z}_1\mathbf{z}_1^T] - \mathbb{E}[\mathbf{z}_1]\mathbb{E}[\mathbf{z}_1^T] \quad (32)$$

$$\mathbf{F}^{new} = (\sum_{n=2}^{\mathrm{T}}\mathbb{E}[\mathbf{z}_n\mathbf{z}_{n-1}^T])(\sum_{n=1}^{\mathrm{T}-1}\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^T])^{-1} \quad (33)$$

$$\Lambda^{new} = \frac{1}{\mathrm{T}-1}\sum_{n=2}^{\mathrm{T}}(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^T] - \mathbf{F}^{new}\mathbb{E}[\mathbf{z}_{n-1}\mathbf{z}_n^T] - \mathbb{E}[\mathbf{z}_n\mathbf{z}_{n-1}^T](\mathbf{F}^{new})^T + \mathbf{F}^{new}\mathbb{E}[\mathbf{z}_n\mathbf{z}_{n-1}^T](\mathbf{F}^{new})^T) \quad (34)$$

$$\mathbf{G}^{new} = (\sum_{n=1}^{\mathrm{T}}\mathbf{x}_n\mathbb{E}[\mathbf{z}_n^T])(\sum_{n=1}^{\mathrm{T}}\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^T])^{-1} \quad (35)$$

$$\Sigma^{new} = \frac{1}{\mathrm{T}}\sum_{n=1}^{\mathrm{T}}(\mathbf{x}_n\mathbf{x}_n^T - \mathbf{G}^{new}\mathbb{E}[\mathbf{z}_n]\mathbf{x}_n^T - \mathbf{x}_n\mathbb{E}[\mathbf{z}_n^T](\mathbf{G}^{new})^T + \mathbf{G}^{new}\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^T](\mathbf{G}^{new})^T) \quad (36)$$

The resulting equations in E-step and M-step of EMN/EMG algorithms are traditionally known as Kalman filtering and Kalman smoothing. We would refer readers to [SS82] for more details.

Appendix B: BoLeRO-HC details

BoLeRO-HC uses Newton's method to iteratively search the optimal solution through the update Eq. 7. The partial gradient is given by:

$$\nabla_{\mathbf{x}_t^M,\eta_t} = \begin{pmatrix} I_t^M\Sigma^{-1}\cdot(\mathbf{x}_t - \mathbf{G}\cdot\mathbb{E}[\mathbf{z}_t]) + 2\sum_{i,j}\delta_B(t,i,j)\eta_{ijt}(\mathbf{x}_t^{(i)} - \mathbf{x}_t^{(j)}) \\ (\|\mathbf{x}_t^{(i)} - \mathbf{x}_t^{(j)}\|^2 - d_{i,j}^2) \quad \text{for all } \delta_B(t,i,j) = 1 \end{pmatrix} \quad (37)$$

where I_t^M is a $(0,1)$-matrix, with each row has exactly one nonzero elements corresponding to the missing marker indices at time t. The auxiliary function $\delta_B(t,i,j)$ is defined as

$$\delta_B(t,i,j) = \begin{cases} 1 & \text{if} \langle i,j,d_{i,j}\rangle \in B \wedge (W_{t,i} = 1 \vee W_{t,j} = 1) \\ 0 & \text{otherwise} \end{cases}$$

By further taking partial derivative, we can get the Hessian,

$$\nabla^2_{\mathbf{x}_t^M,\eta_t} = \frac{\partial\nabla_{\mathbf{x}_t^M,\eta_t}}{\partial(\mathbf{x}_t^M,\eta_t)^T} \quad (38)$$

Details are straightforward and omitted for brevity.