Gaussian Process II:
Learning hyper parameters

Choosing kernel parameters

$$k(t_i, t_j) = \sigma_f^2 \exp\left\{-\frac{1}{2\ell^2}(t_i - t_j)^2\right\}$$

- Choose some *hyperparameters*: $\sigma_f = 14$, $\ell = 50$

$$t = \begin{bmatrix} 0700 \\ 0800 \\ 1029 \end{bmatrix} \qquad K(t,t) = \{k(t_i, t_j)\}_{i,j} = \begin{bmatrix} 196 & 26.5 & 00.0 \\ 26.5 & 196 & 0.01 \\ 00.0 & 0.01 & 196 \end{bmatrix}$$

- Choose some *hyperparameters*: $\sigma_f = 7$, $\ell = 100$

$$t = \begin{bmatrix} 0700 \\ 0800 \\ 1029 \end{bmatrix} \qquad K(t,t) = \{k(t_i, t_j)\}_{i,j} = \begin{bmatrix} 49.0 & 29.7 & 00.2 \\ 29.7 & 49.0 & 03.6 \\ 00.2 & 03.6 & 49.0 \end{bmatrix}$$

- Choose some *hyperparameters*: $\sigma_f = 7$, $\ell = 500$

$$t = \begin{bmatrix} 0700 \\ 0800 \\ 1029 \end{bmatrix} \qquad K(t,t) = \{k(t_i, t_j)\}_{i,j} = \begin{bmatrix} 49.0 & 48.0 & 39.5 \\ 48.0 & 49.0 & 44.1 \\ 39.5 & 44.1 & 49.0 \end{bmatrix}$$

Sampling from a Covariance matrix

Scalar case

$$x \sim \mathcal{N}(\mu, \sigma^2)$$

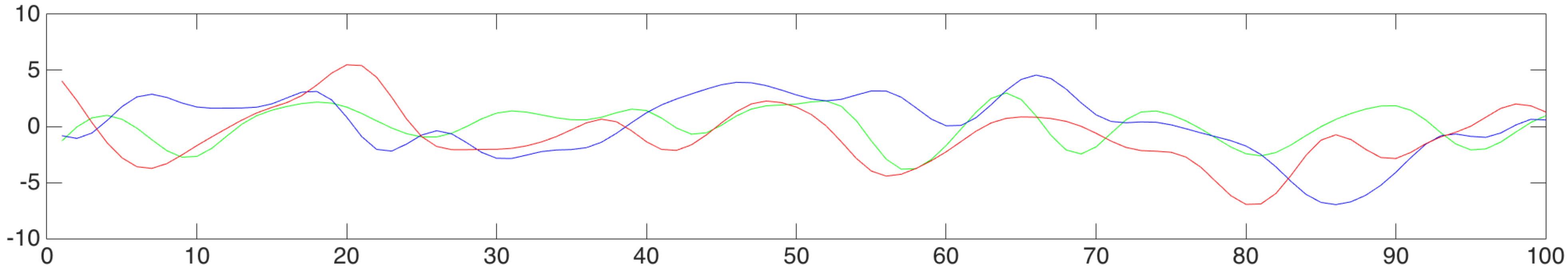
$$x \sim \mu + \sigma(\mathcal{N}(0, 1)$$

Vector case in GP setting

$$f_* \sim \mu + B(\mathcal{N}(0, I)$$

$$BB^T = \Sigma_*$$

Cholesky decomposition accomplishes this factorization
where are B and B^T upper and lower triangular matrices



```
function [ Km ] = GaussianF( numPts,s,l )
%Priors for GP ; numPts=100, s =2, l=3

Km = zeros(numPts,numPts);
for i = 1: numPts
    for j = 1: numPts
        Km(i,j) = (s^2)*exp(-(0.5/l^2)*(i-j)^2);
    end
end
end
```

```
F=chol(Km)'*randn(100,1);
```

If f and y are jointly Gaussian:

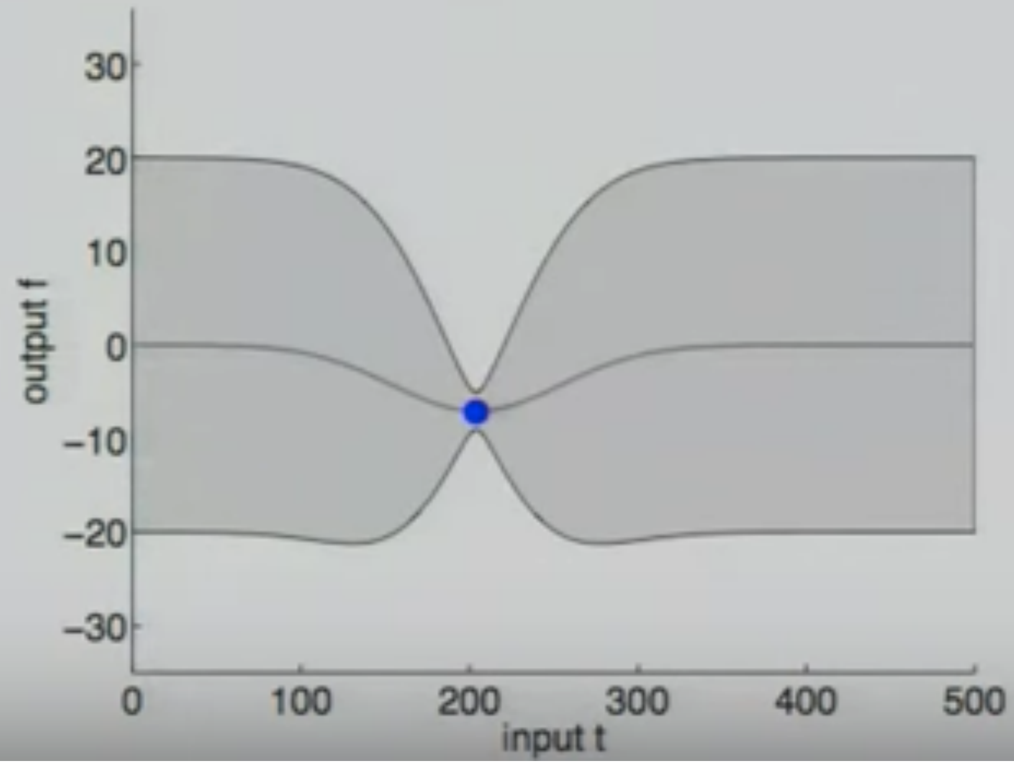
$$\begin{bmatrix} f \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_f \\ m_y \end{bmatrix}, \begin{bmatrix} K_{ff} & K_{fy} \\ K_{fy}^T & K_{yy} \end{bmatrix} \right)$$

Then:

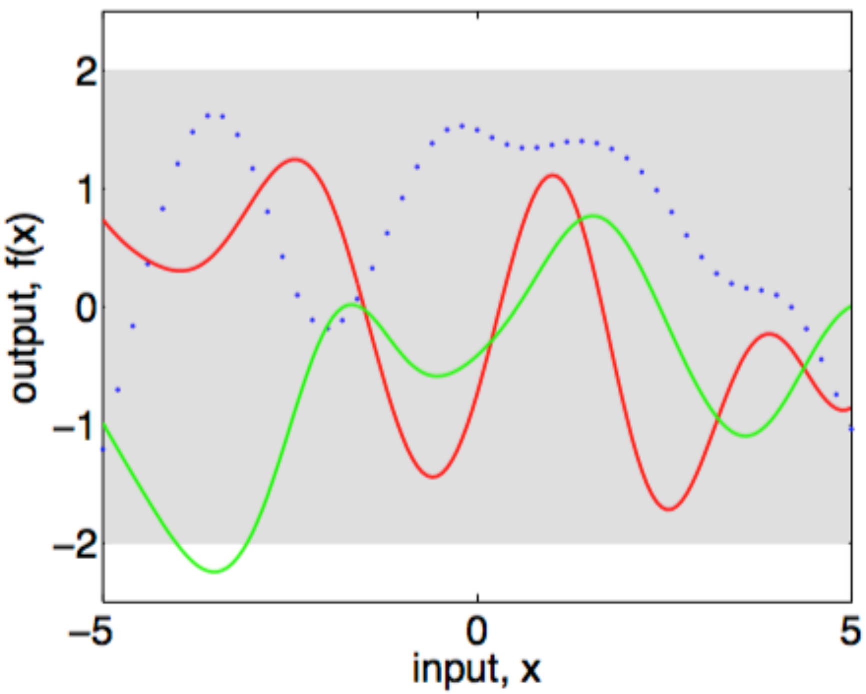
$$f|y \sim \mathcal{N}\left(K_{fy}K_{yy}^{-1}(y - m_y) + m_f \ , \ K_{ff} - K_{fy}K_{yy}^{-1}K_{fy}^T\right)$$

- Use conditioning to update the posterior:

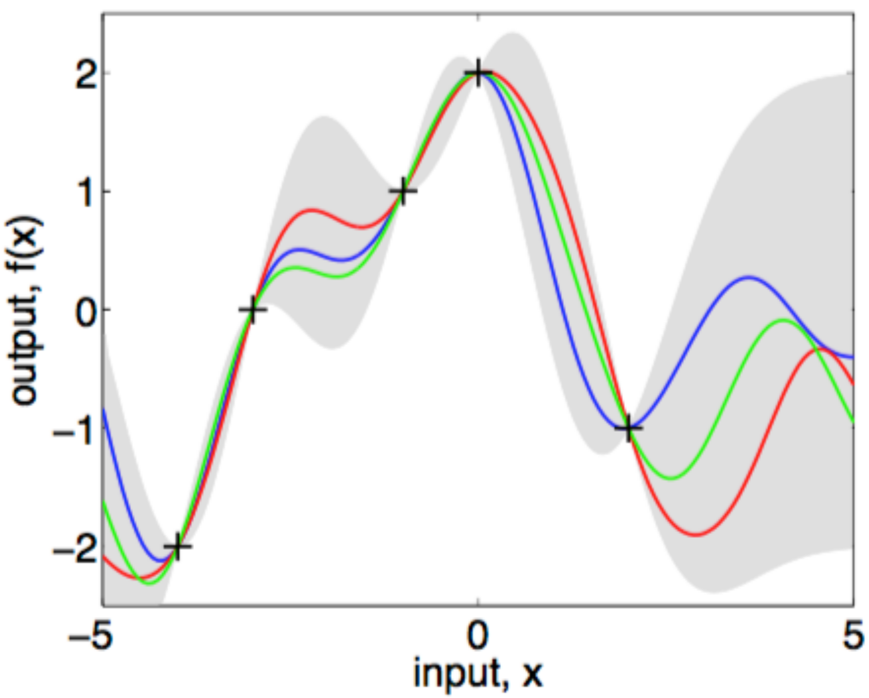
$$f|y(204) \sim \mathcal{N}\left(K_{fy}K_{yy}^{-1}(y(204) - m_y)\ ,\ K_{ff} - K_{fy}K_{yy}^{-1}K_{fy}^{T}\right)$$



An impractical but intuitively appealing way to think about
Bayesian reasoning in the GP in the noiseless case:
Of all the possible functions, just keep the ones that fit the data

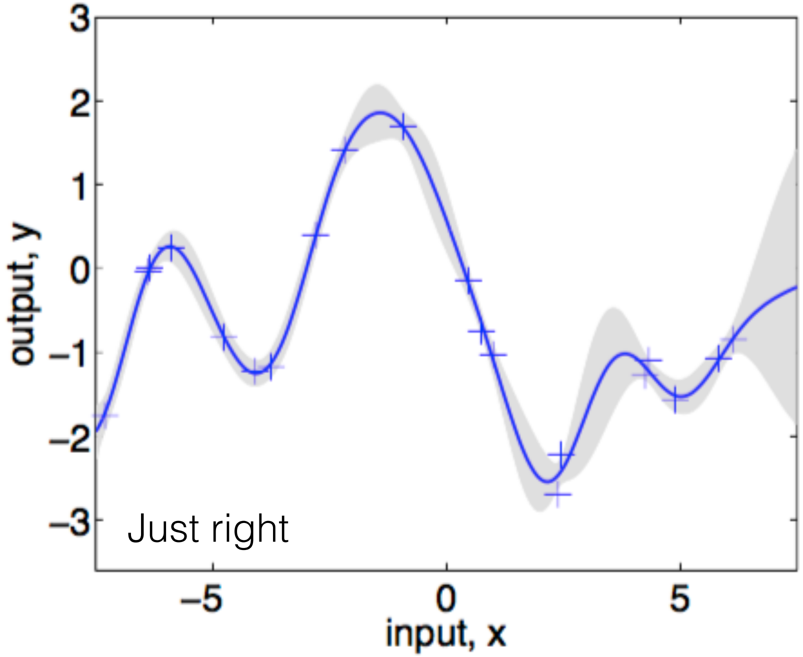


(a), prior

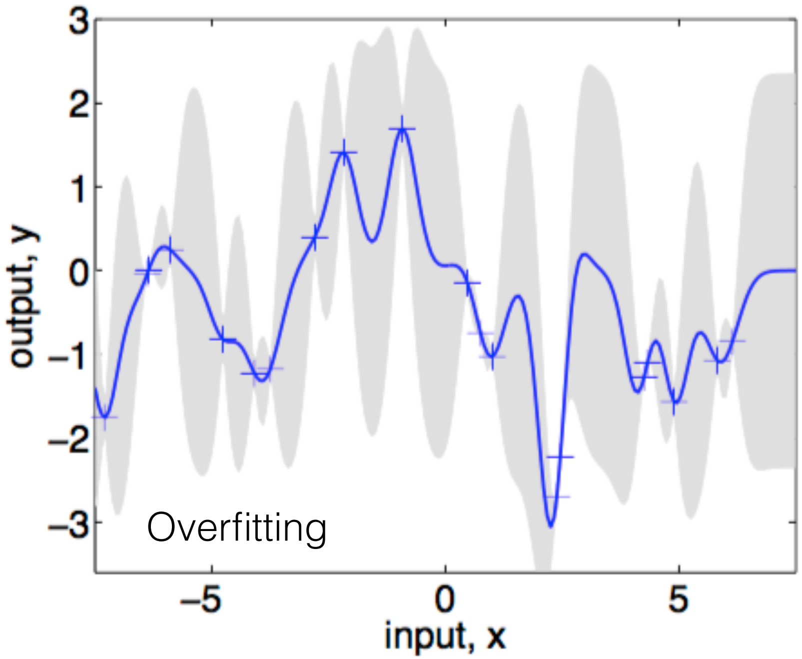


(b), posterior

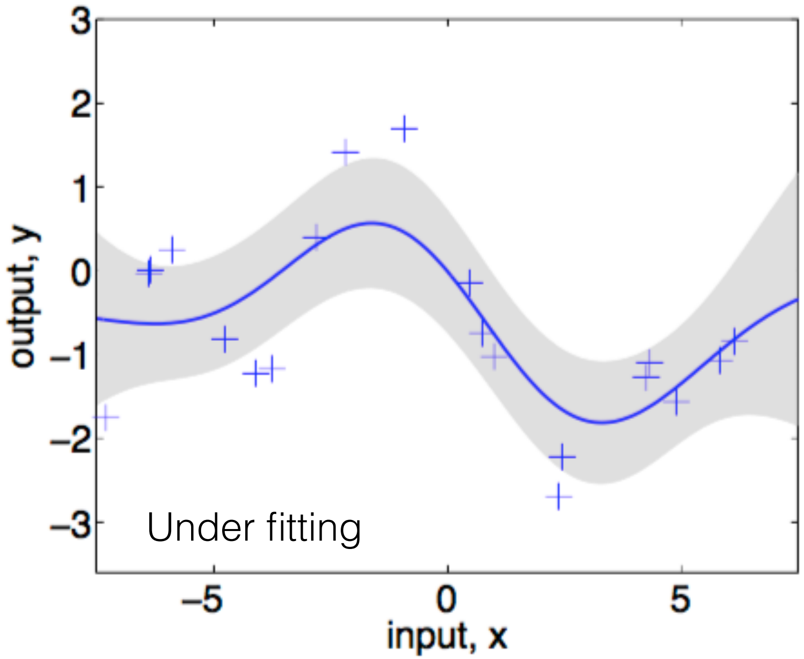
A practical problem: choosing the best set of hyperparameters



(a), $\ell = 1$



(b), $\ell = 0.3$



(c), $\ell = 3$

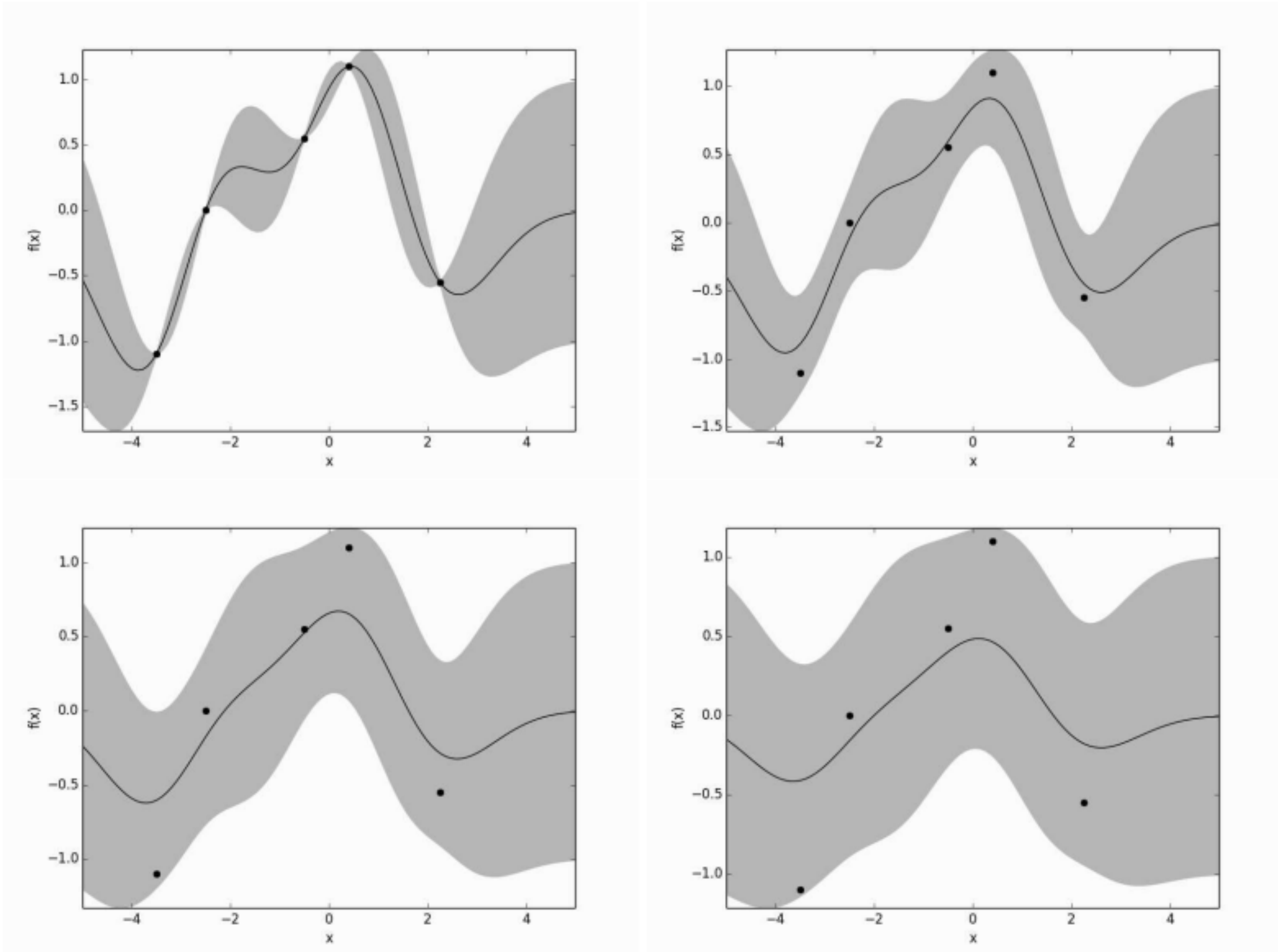


FIGURE 18.4 The effects of adding noise to the estimate of the covariance in the training data with the squared exponential kernel. Each plot shows the mean and 2 standard deviation error bars for a Gaussian process fitted to the five datapoints marked with dots. *Top left:* $\sigma_n = 0.0$, *top right:* $\sigma_n = 0.2$, *bottom left:* $\sigma_n = 0.4$, *bottom right:* $\sigma_n = 0.6$

f

y

$$p(y, f) = \mathcal{N}\left(\begin{bmatrix} f \\ y \end{bmatrix} ; \begin{bmatrix} m_f \\ m_y \end{bmatrix} , \begin{bmatrix} K_{ff} & K_{fy} \\ K_{fy}^T & K_{yy} \end{bmatrix} \right)$$

Conditioning

$$f|y \sim \mathcal{N}\left(K_{fy}^T K_{yy}^{-1} (y - m_y) + m_f , \; K_{ff} - K_{fy}^T K_{yy}^{-1} K_{fy} \right)$$

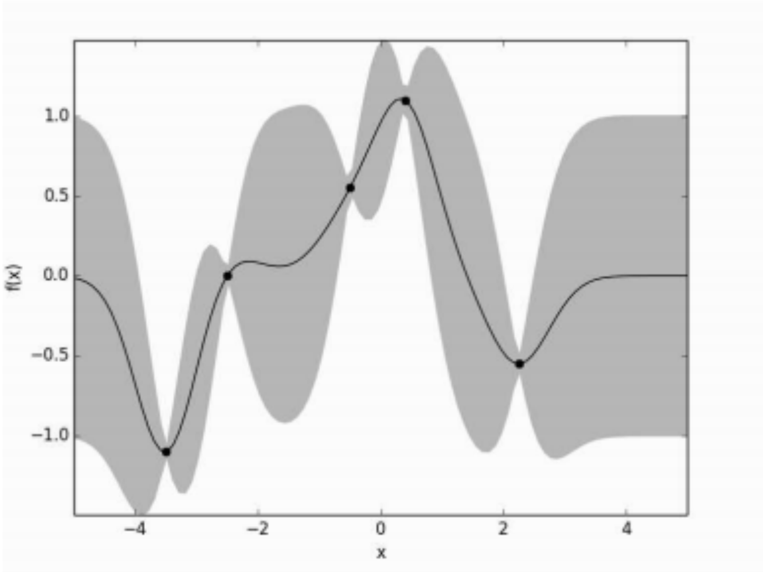
---

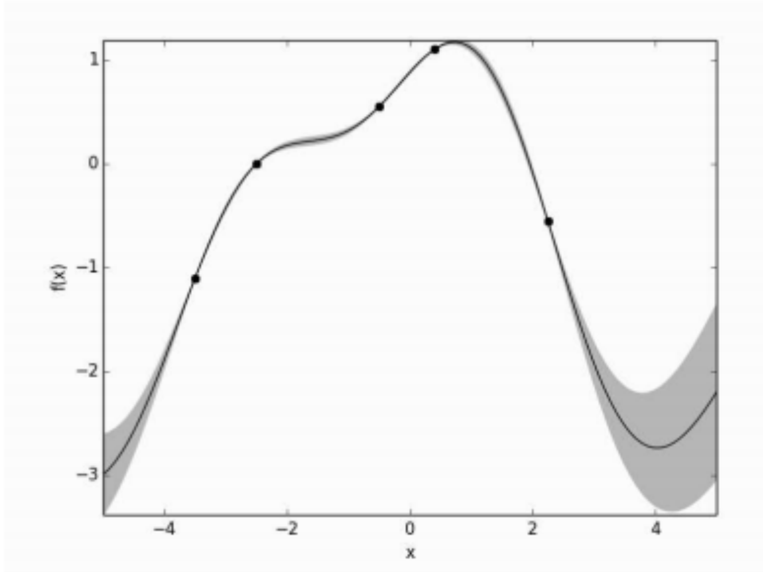
3 pts

$$\begin{bmatrix} f \\ y_* \end{bmatrix} \sim \mathcal{N}\left(0 , \begin{bmatrix} \underbrace{\begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{33} \\ K_{31} & K_{32} & K_{33} \end{bmatrix}}_{K} & \underbrace{\begin{bmatrix} K_{1*} \\ K_{2*} \\ K_{3*} \end{bmatrix}}_{K_*} \\ \underbrace{\begin{bmatrix} K_{*1} & K_{*2} & K_{*3} \end{bmatrix}}_{K_*^T} & \underbrace{\begin{bmatrix} K_{**} \end{bmatrix}}_{K_{**}} \end{bmatrix} \right)$$

$$m^* = K_*^T K^{-1} y$$

$$\sigma^* = K_{**} - K_*^T K^{-1} K_*$$





l=0.5

Xs are relatively uncorrelated

l=2.0

Xs are are more uncorrelated

Now for learning hyper parameters …

$$P(\mathbf{f}|\mathbf{X}) = (2\pi)^{\frac{n}{2}}|\mathbf{K}|^{\frac{1}{2}}\exp(-\frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f})$$

$\mathbf{f}|\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$

$$\log p(\mathbf{f}|X) = -\tfrac{1}{2}\mathbf{f}^\top K^{-1}\mathbf{f} - \tfrac{1}{2}\log|K| - \tfrac{n}{2}\log 2\pi$$

$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, K + \sigma_n^2 I)$

$$\log p(\mathbf{y}|X) = -\tfrac{1}{2}\mathbf{y}^\top(K + \sigma_n^2 I)^{-1}\mathbf{y} - \tfrac{1}{2}\log|K + \sigma_n^2 I| - \tfrac{n}{2}\log 2\pi$$

Marsland's notation $\mathbf{y}$ -> $\mathbf{t}$

$$\log P(\mathbf{t}|\mathbf{x},\boldsymbol{\theta}) = -\frac{1}{2}\mathbf{t}^T(\mathbf{K}+\sigma_n^2\mathbf{I})^{-1}\mathbf{t} - \frac{1}{2}\log|\mathbf{K}+\sigma_n^2\mathbf{I}| - \frac{N}{2}\log 2\pi$$

$$\mathbf{Q} = (\mathbf{K}+\sigma_n^2\mathbf{I})$$

$$\begin{aligned} \frac{\partial \mathbf{Q}^{-1}}{\partial \boldsymbol{\theta}} &= -\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \boldsymbol{\theta}}\mathbf{Q}^{-1} \\ \frac{\partial \log|\mathbf{Q}|}{\partial \boldsymbol{\theta}} &= \mathrm{trace}\left(\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \boldsymbol{\theta}}\right) \end{aligned}$$

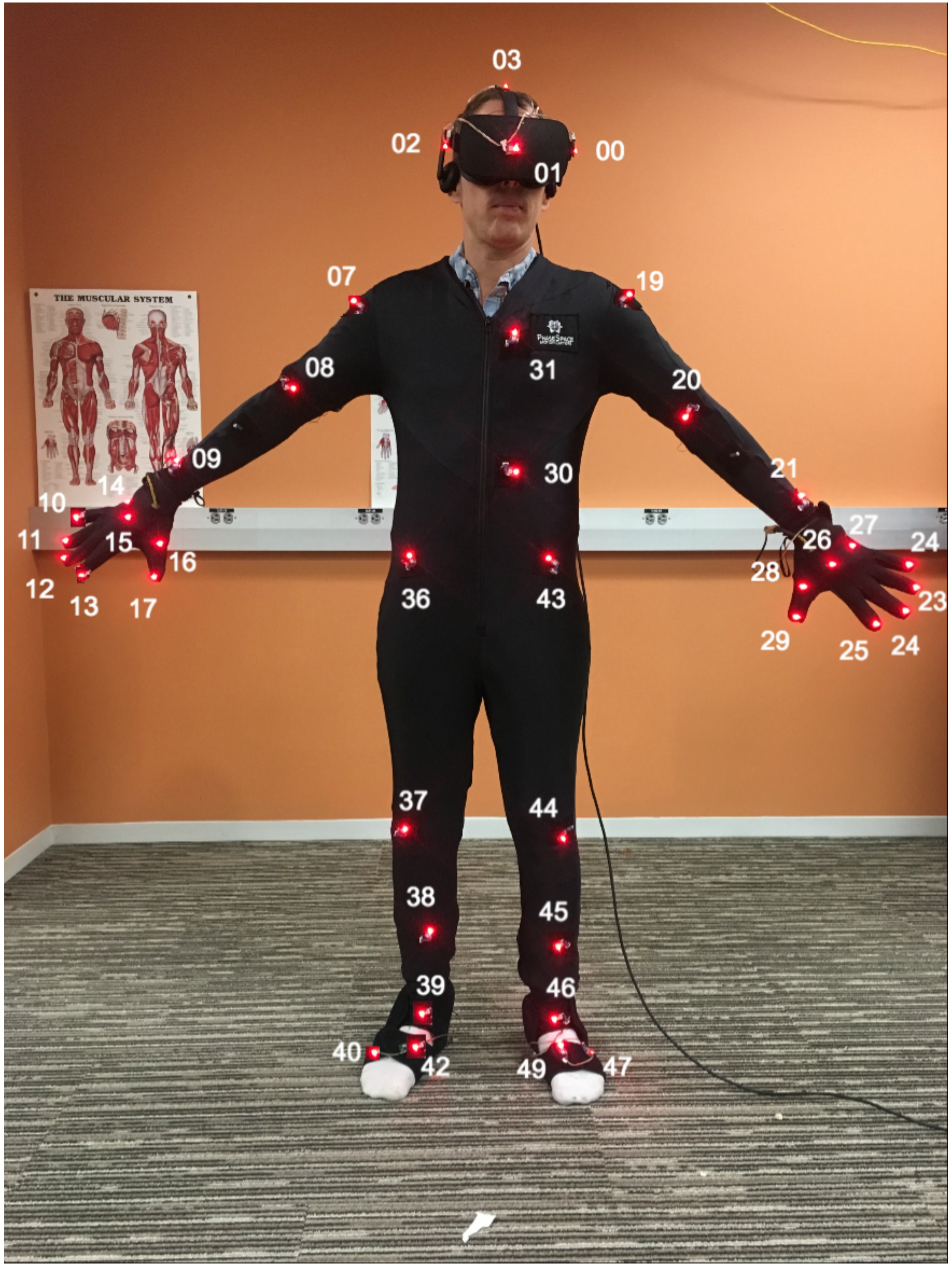
$$\begin{aligned} k(\mathbf{x}, \mathbf{x}') &= \exp(\sigma_f) \exp\left(-\frac{1}{2} \exp(\sigma_l) |\mathbf{x} - \mathbf{x}'|^2\right) + \exp(\sigma_n)\mathbf{I} \\ &= k' + \exp(\sigma_n)\mathbf{I} \end{aligned}$$

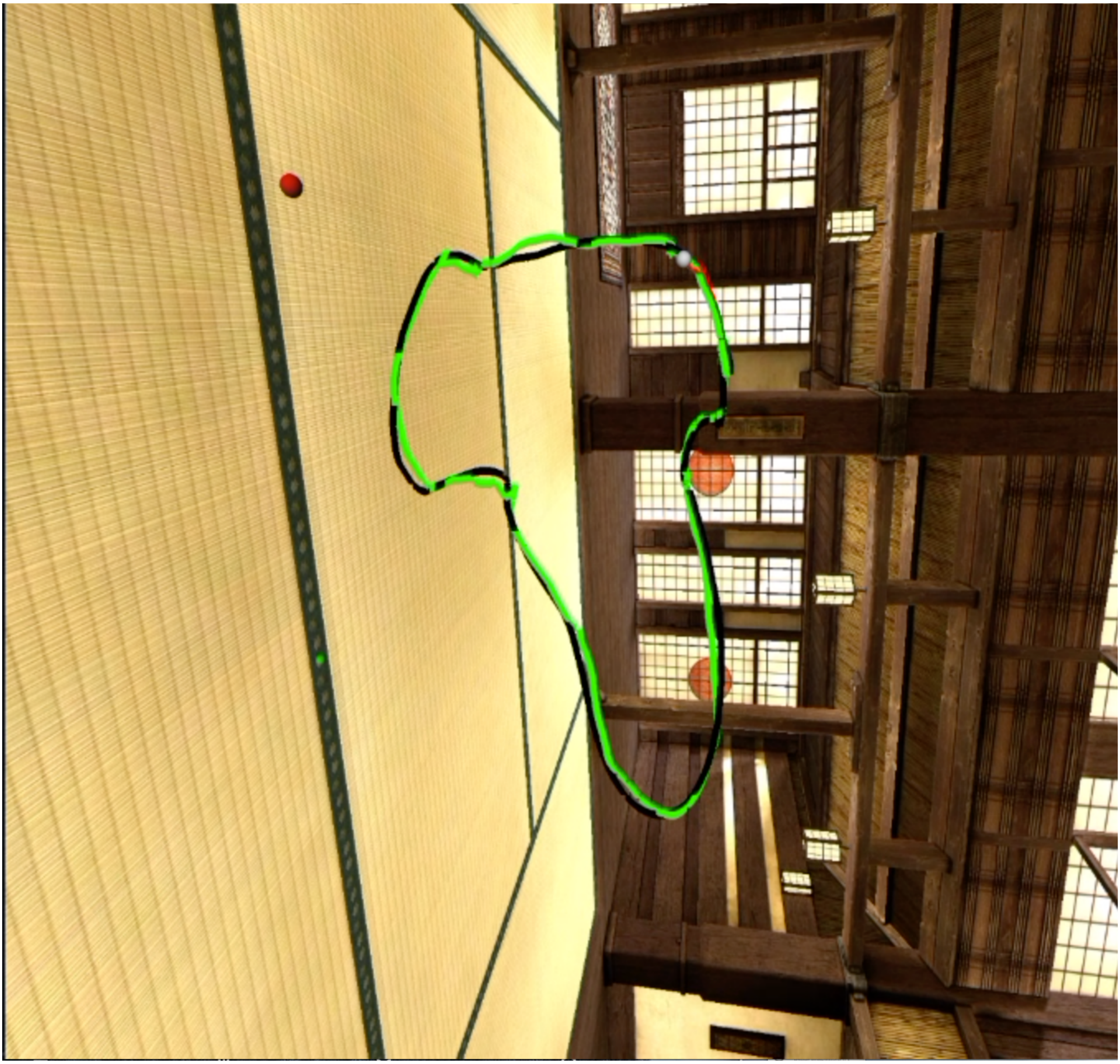
(Use exp(parameter) to make sure they are positive

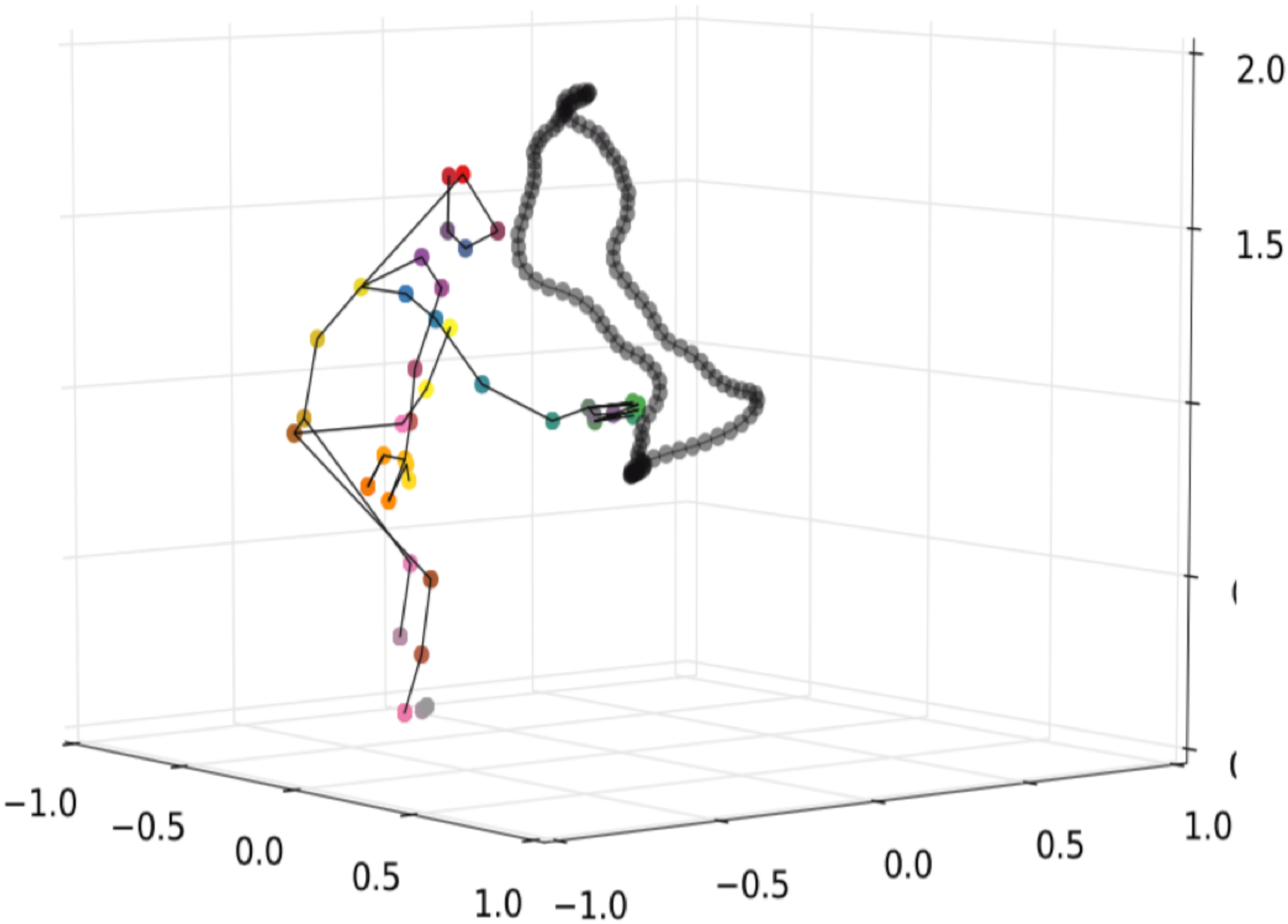
$$\frac{\partial k}{\partial \sigma_f} = k'$$

$$\frac{\partial k}{\partial \sigma l} = k' \times \left(-\frac{1}{2} \exp(\sigma_l) |\mathbf{x} - \mathbf{x}'|^2\right)$$

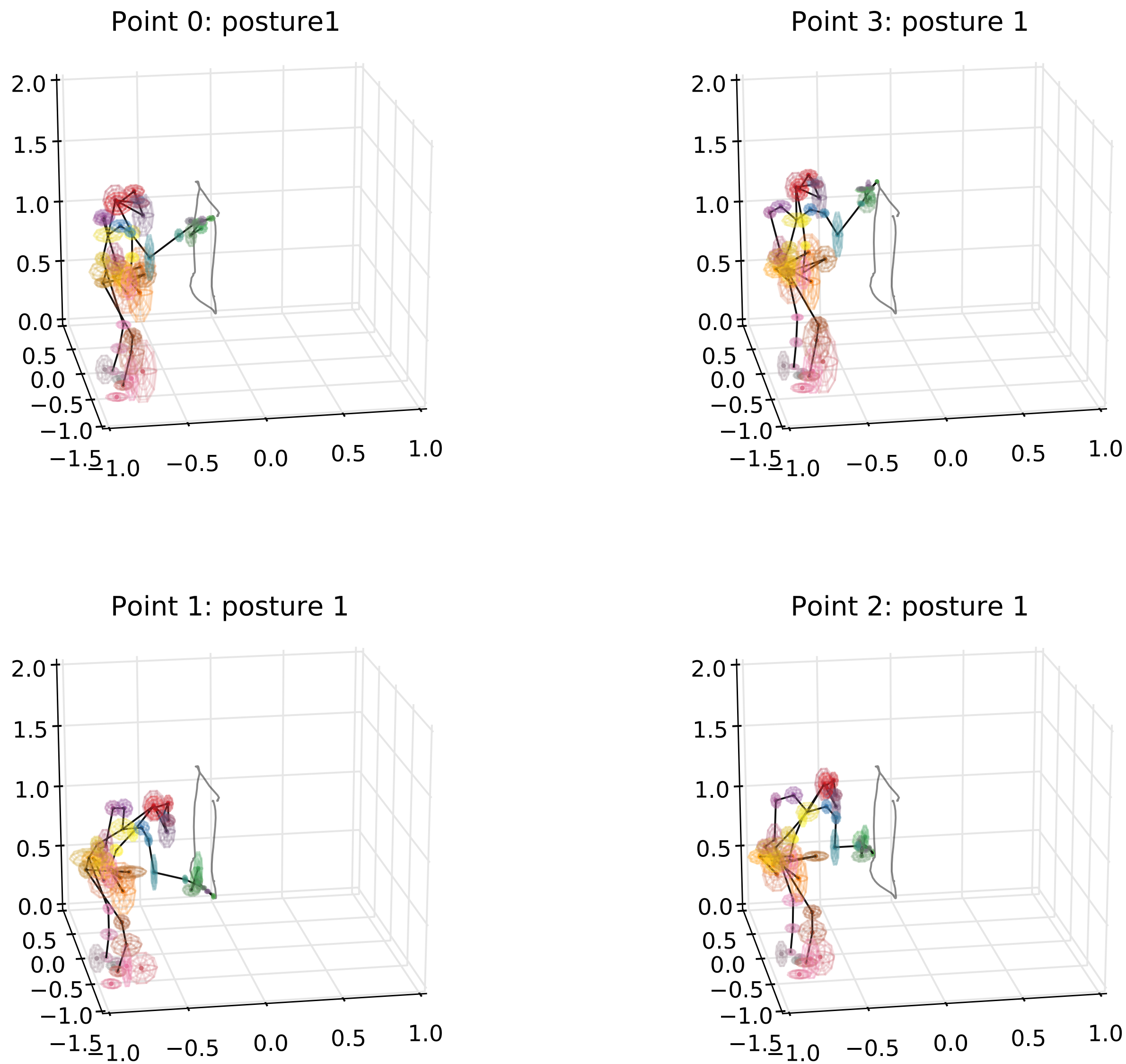
$$\frac{\partial k}{\partial \sigma_n} = \exp(\sigma_n)\mathbf{I}$$


03
02
00
01
07
19
THE MUSCULAR SYSTEM
08
31
20
09
30
21
14
10
27
11
15
26
24
16
28
12
13
17
36
43
23
29
25
24
37
44
38
45
39
46
40
42
49
47

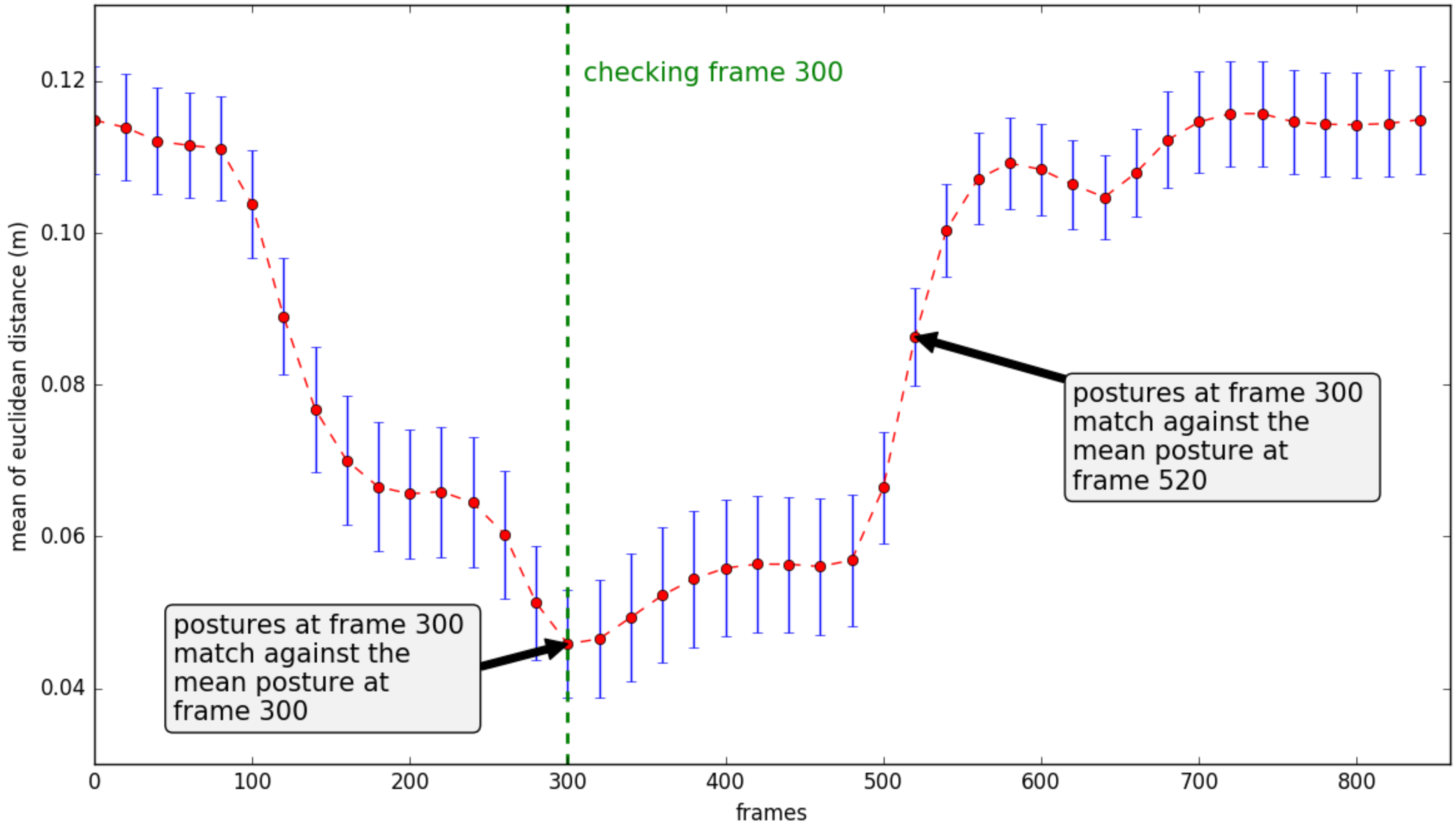


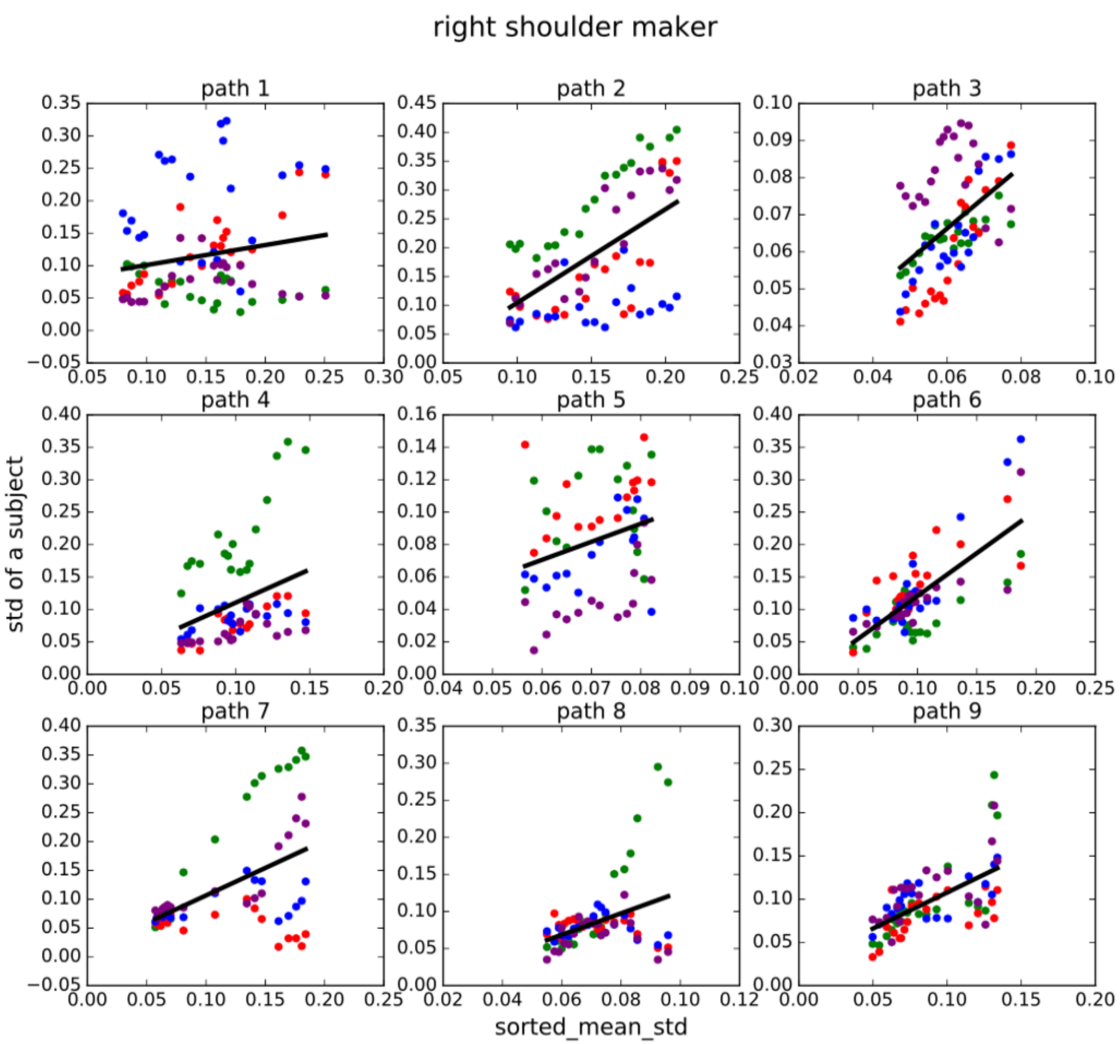


−1.0
−0.5
0.0
0.5
1.0
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

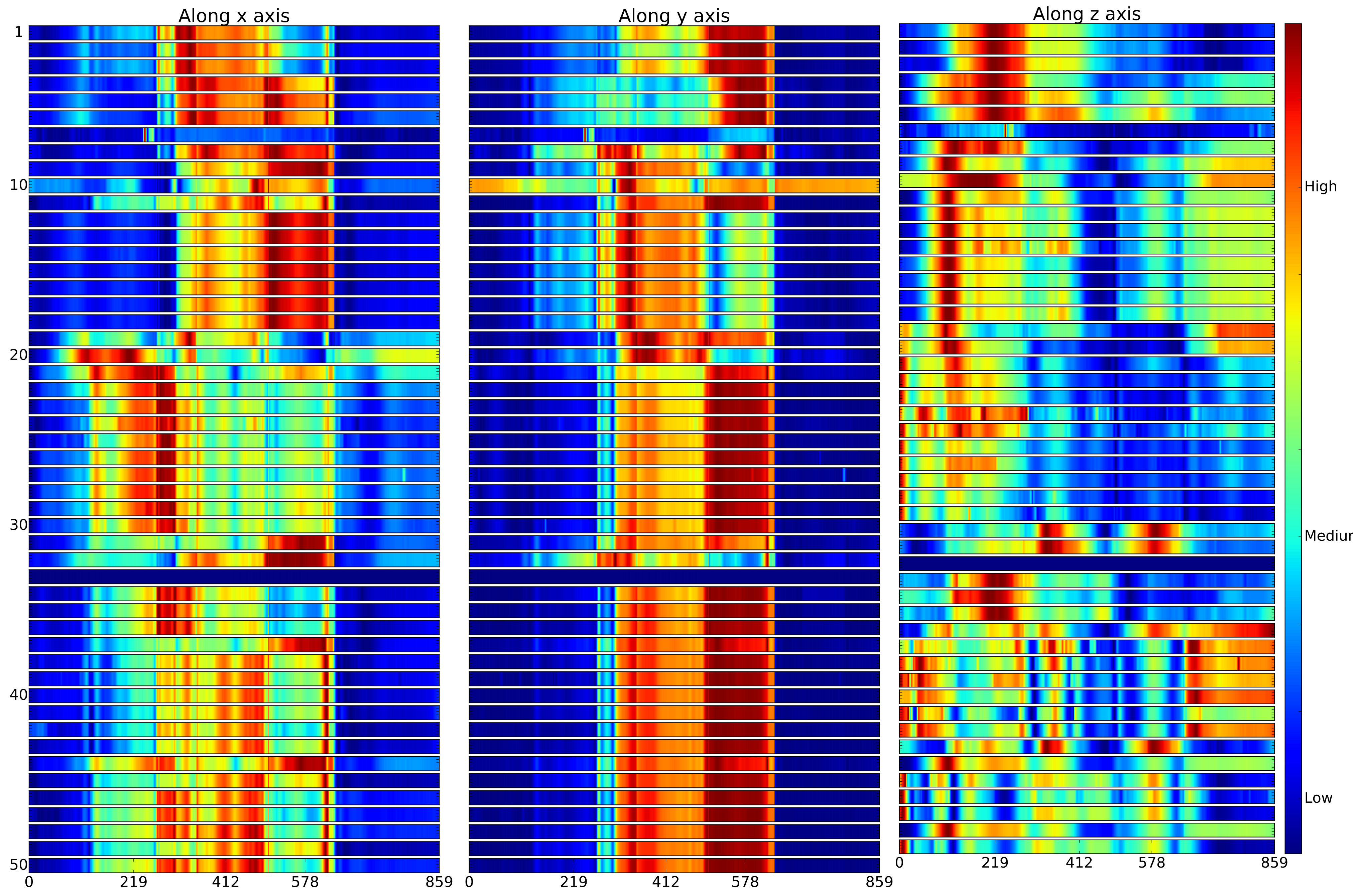

Point 0: posture1
Point 3: posture 1
Point 1: posture 1
Point 2: posture 1

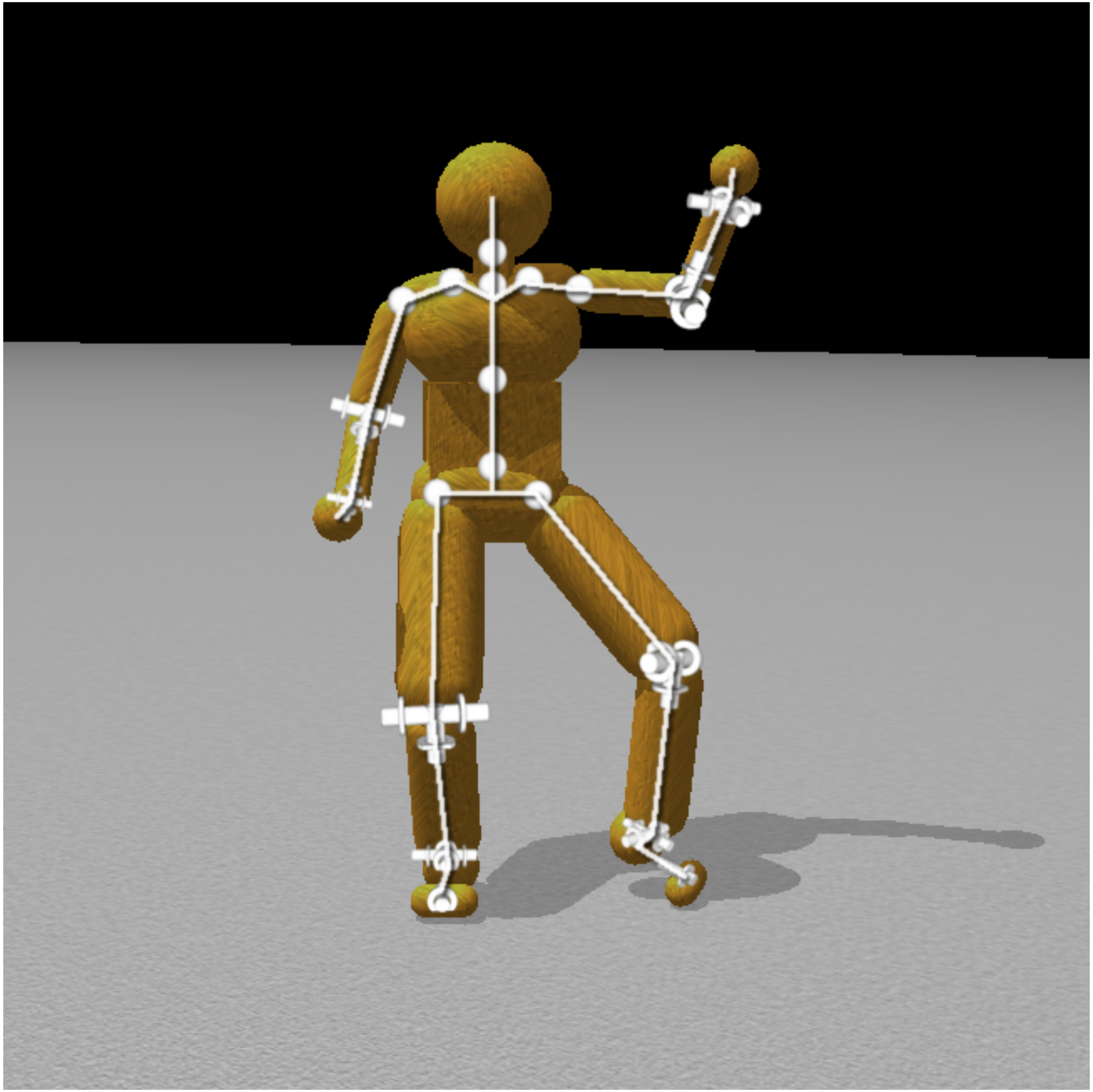
The postures at frame 300 do not kook like postures at any other frame





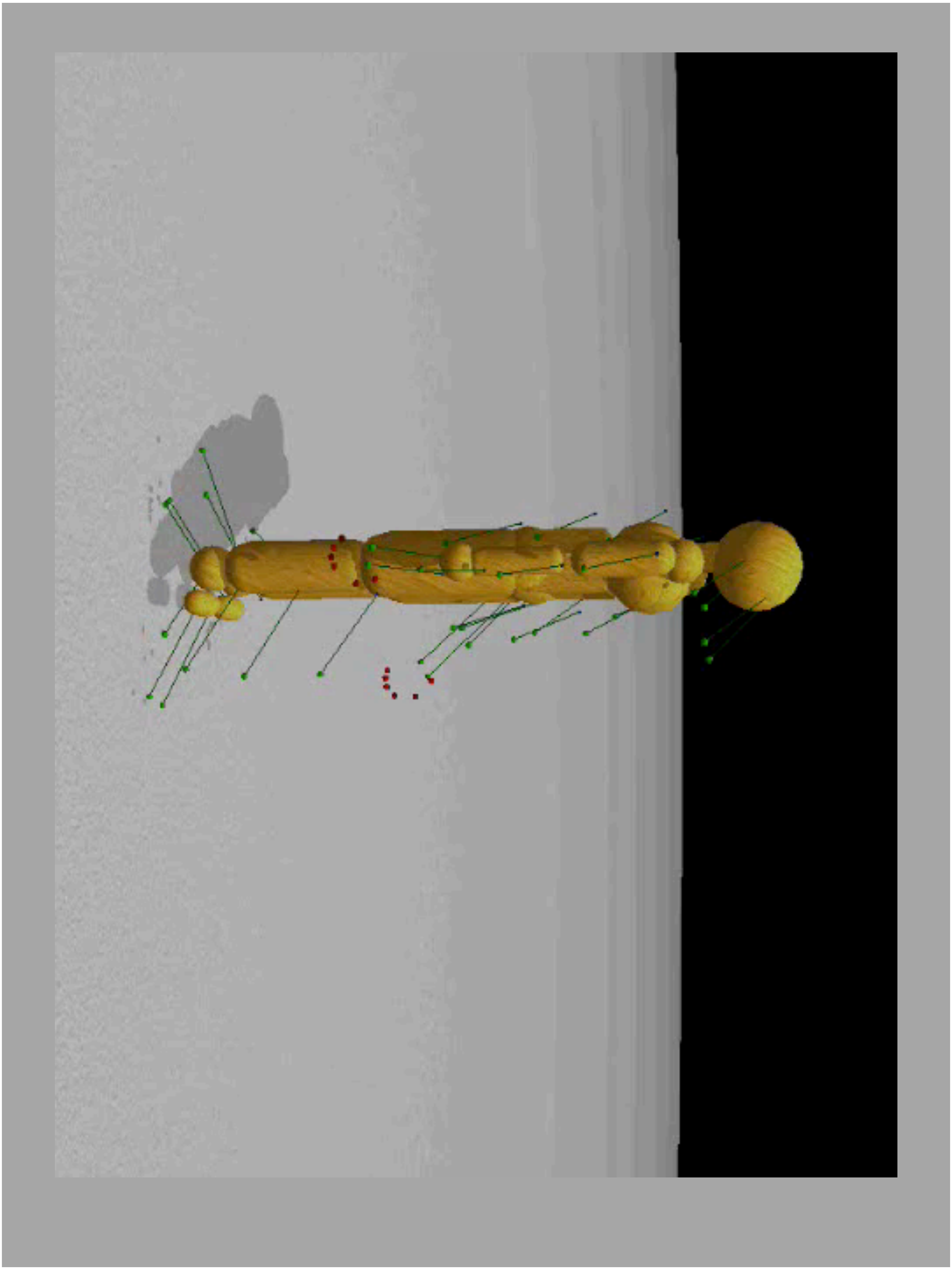
R values	right_index_figure	left_index_figure	right shoulder	head
path1	0.87	0.53	0.5	0.4
path2	0.77	0.83	0.77	0.76
path3	0.76	0.48	0.75	0.57
path4	0.74	0.67	0.68	0.63
path5	0.76	0.67	0.49	0.62
path6	0.81	0.66	0.82	0.71
path7	0.7	0.64	0.75	0.8
path8	0.85	0.66	0.52	0.63
path9	0.8	0.8	0.68	0.64

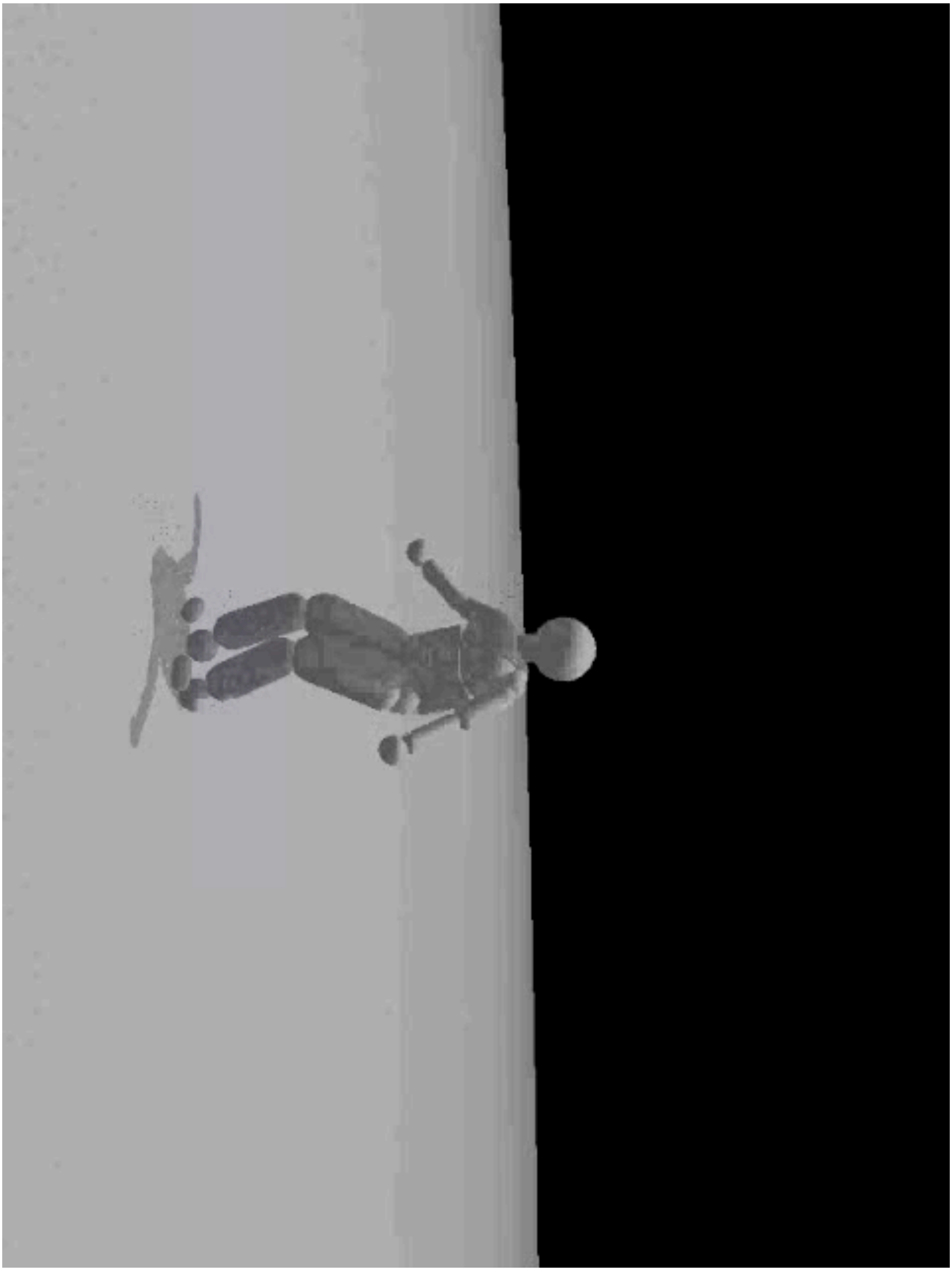

Along x axis
Along y axis
Along z axis
1
10
20
30
40
50
0
219
412
578
859
High
Mediur
Low

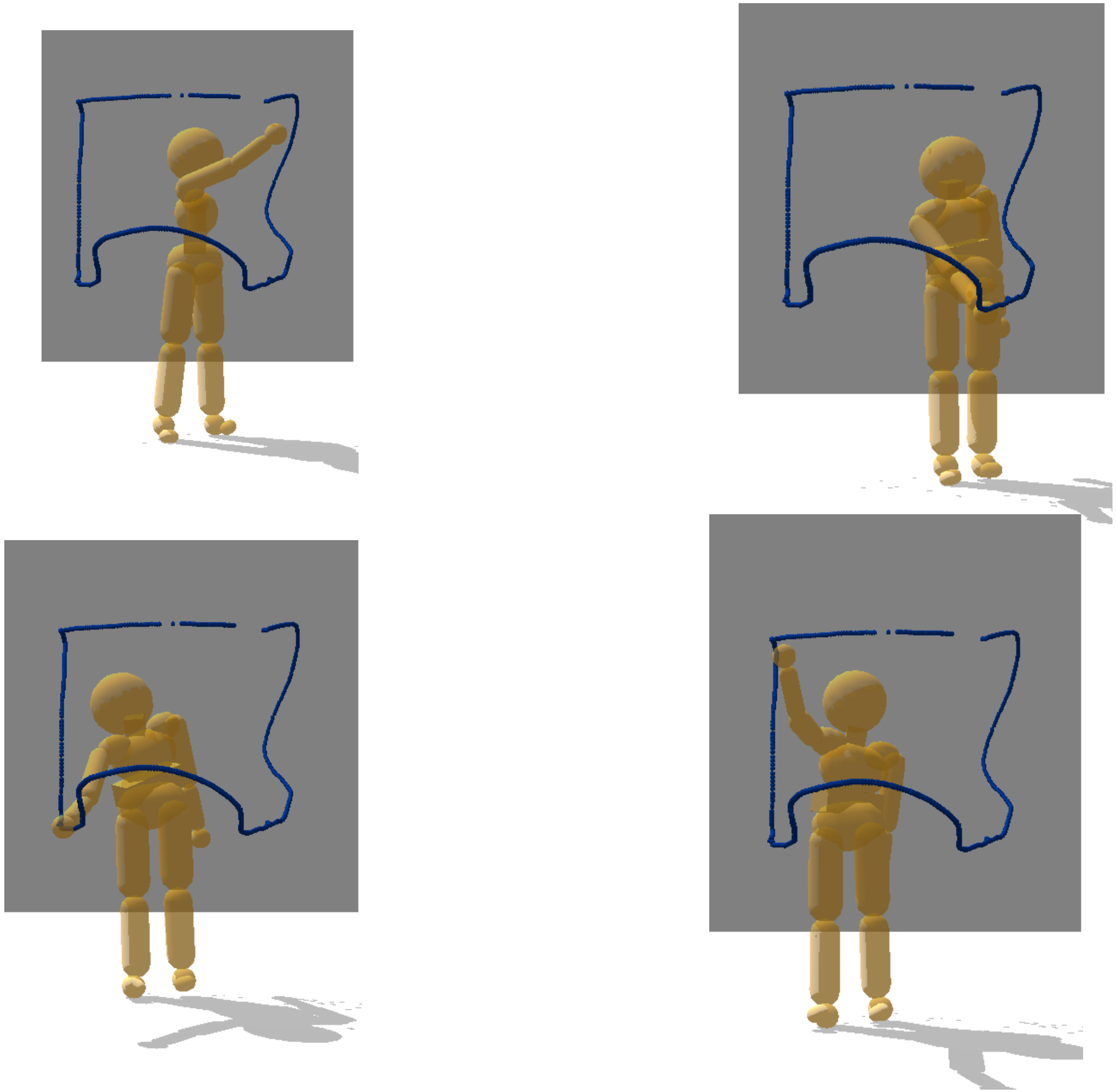


Dynamic Model

Including
50 DOFs,
Lengths,
Masses,
Inertial Tensors







A subject traces the square curve