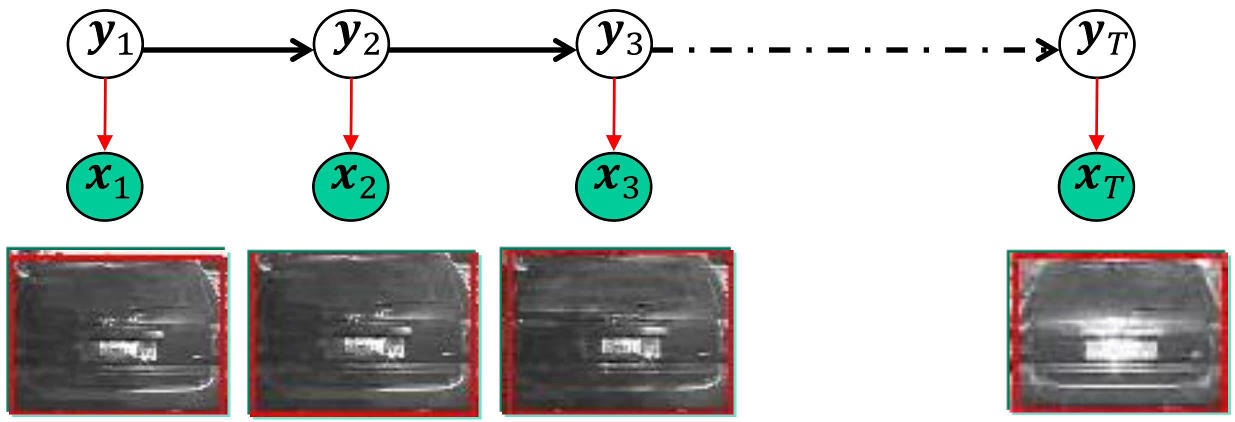
Linear Dynamical Systems (Kalman Filters)

(a) Filtering and Smoothing in LDS
(b) EM in LDS

Linear Dynamical Systems (LDS)



$$\boldsymbol{x}_t = \mathbf{W}\boldsymbol{y}_t + \boldsymbol{e}_t \qquad \mathbf{e} \sim N(\boldsymbol{e}|\mathbf{0}, \boldsymbol{\Sigma})$$

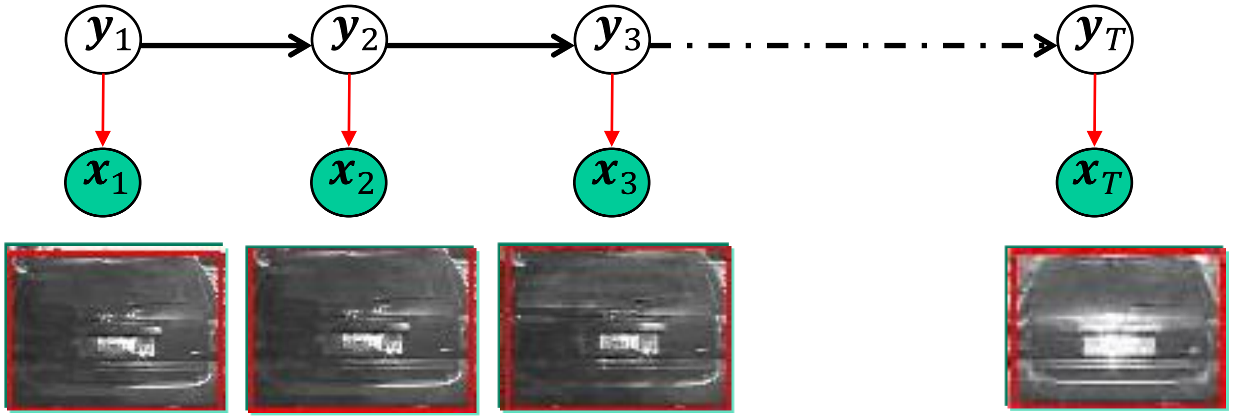
Transition model

$$\boldsymbol{y}_1 = \boldsymbol{\mu}_0 + \boldsymbol{u} \qquad \boldsymbol{u} \sim N(\boldsymbol{u}|\mathbf{0}, \boldsymbol{P}_0)$$

$$\boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{y}_{t-1} + \boldsymbol{v}_t \qquad \boldsymbol{v} \sim N(\boldsymbol{v}|\mathbf{0}, \boldsymbol{\Gamma})$$

Parameters: $\theta = \{\boldsymbol{W}, \boldsymbol{A}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{P}_0\}$

Linear Dynamical Systems (LDS)



First timestamp: $p(\boldsymbol{y}_1) = N(\boldsymbol{y}_1 | \boldsymbol{\mu}_0, \boldsymbol{P}_0)$

Transition Probability : $p(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}) = N(\boldsymbol{y}_t | \boldsymbol{A}\boldsymbol{y}_{t-1}, \boldsymbol{\Gamma})$

Emission: $p(\boldsymbol{x}_t | \boldsymbol{y}_t) = N(\boldsymbol{x}_t | \boldsymbol{W}\boldsymbol{y}_t, \boldsymbol{\Sigma})$

HMM vs LDS

HMM	LDS
Markov Chain with discrete latent variables	Markov Chain with continuous latent variables

HMM

Markov Chain

with discrete latent variables

$p(\boldsymbol{y_1})$ $\quad \boldsymbol{\pi} \quad Kx1$

$p(\boldsymbol{y_t}|\boldsymbol{y_{t-1}})$ $\quad \boldsymbol{A} \quad KxK$

$p(\boldsymbol{x_t}|\boldsymbol{y_t})$ $\quad \boldsymbol{B} \quad LxK$

or

$p(\boldsymbol{x_t}|\boldsymbol{y_t})$ $\quad K$ distributions

LDS

Markov Chain

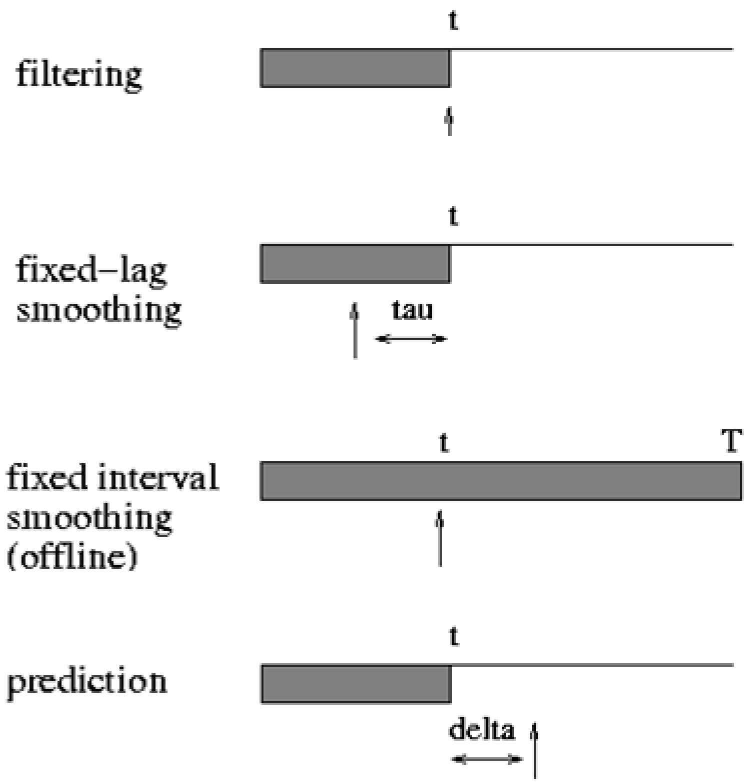
with continuous latent variables

$$p(\boldsymbol{y}_1) = N(\boldsymbol{y}_1|\boldsymbol{\mu}_0, \boldsymbol{P}_0)$$

$$p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}) = N(\boldsymbol{y}_t|\boldsymbol{A}\boldsymbol{y}_{t-1}, \boldsymbol{\Gamma})$$

$$p(\boldsymbol{x}_t|\boldsymbol{y}_t) = N(\boldsymbol{x}_t|\boldsymbol{W}\boldsymbol{y}_t, \boldsymbol{\Sigma})$$

LDS



$$p(\boldsymbol{y}_t | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$$

$$p(\boldsymbol{y}_{t-\tau} | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{\boldsymbol{t}})$$

$$p(\boldsymbol{y}_t | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T)$$

$$p(\boldsymbol{y}_{t+\delta} | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{\boldsymbol{t}})$$

$$p(\boldsymbol{x}_{t+\delta} | \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{\boldsymbol{t}})$$

Filtering

Filtering: $\hat{a}(\boldsymbol{y}_t) = p(\boldsymbol{y}_t|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_t)$

The filtered probability is a Gaussian: $\hat{a}(\boldsymbol{y}_t) = N(\boldsymbol{y}_t|\boldsymbol{\mu_t}, \boldsymbol{V_t})$

Hence we need to recursively compute: $\boldsymbol{\mu_t}, \boldsymbol{V_t}$

HMM $$c_t\hat{a}(\boldsymbol{z}_t) = p(\boldsymbol{x}_t|\boldsymbol{z}_t)\sum_{\boldsymbol{z}_{t-1}} \hat{a}(\boldsymbol{z}_{t-1})p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1})$$

$$c_t\hat{a}(\boldsymbol{y}_t) = p(\boldsymbol{x}_t|\boldsymbol{y}_t)\int_{\boldsymbol{y}_{t-1}} \hat{a}(\boldsymbol{y}_{t-1})p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1})\mathrm{d}\boldsymbol{y}_{t-1}$$

Filtering

$$\int_{\boldsymbol{y}_{t-1}} \hat{a}(\boldsymbol{y}_{t-1}) p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}) \mathrm{d}\boldsymbol{y}_{t-1}$$

$$= \int N(\boldsymbol{y}_t|\boldsymbol{A}\boldsymbol{y}_{t-1}, \boldsymbol{\Gamma}) N(\boldsymbol{y}_{t-1}|\boldsymbol{\mu}_{t-1}, \boldsymbol{V}_{t-1}) \mathrm{d}\boldsymbol{y}_{t-1}$$

$$= N(\boldsymbol{y}_t|\boldsymbol{A}\boldsymbol{\mu}_{t-1}, \mathbf{P}_{t-1})$$

Using the technique “completing the square

$$\boldsymbol{P}_{t-1} = \boldsymbol{A}\boldsymbol{V}_{t-1}\boldsymbol{A}^T + \boldsymbol{\Gamma}$$

Filtering

$$c_t \hat{a}(\boldsymbol{y}_t) = N(\boldsymbol{x}_t | W\boldsymbol{y}_t, \boldsymbol{\Sigma}) N(\boldsymbol{y}_t | \boldsymbol{A}\boldsymbol{y}_{t-1}, \mathbf{P}_{t-1})$$

$$\boldsymbol{c_t} N(\boldsymbol{y}_t | \boldsymbol{\mu_t}, \boldsymbol{V_t}) = N(\boldsymbol{x}_n | \boldsymbol{W}\boldsymbol{y}_t, \boldsymbol{\Sigma}) N(\boldsymbol{y}_t | \boldsymbol{A}\boldsymbol{y}_{t-1}, \mathbf{P}_{t-1})$$

Which gives the updates:

$$\boldsymbol{\mu_t} = \boldsymbol{A\mu_{t-1}} + \boldsymbol{K_t}(\boldsymbol{x_t} - \boldsymbol{WA\mu_{t-1}})$$

$$\boldsymbol{V_t} = (\boldsymbol{I} - \boldsymbol{K_t W})\boldsymbol{P_{t-1}}$$

$$\mathbf{K}_{\boldsymbol{t}} = \boldsymbol{P_{t-1} W^T} (\boldsymbol{W P_{t-1} W^T} + \boldsymbol{\Sigma})^{-1}$$

Kalman Gain

and:

$$c_t = \boldsymbol{N}(\boldsymbol{x_t} | \boldsymbol{WA\mu_{t-1}}, \boldsymbol{WP_{t-1}W^T} + \boldsymbol{\Sigma})$$

Filtering

Start of the recursion

$$c_1 \hat{a}(\boldsymbol{y}_1) = p(\boldsymbol{y}_1)p(\boldsymbol{x}_1|\boldsymbol{y}_1)$$
$$c_1 N(\boldsymbol{y}_1|\boldsymbol{\mu}_1, \boldsymbol{V}_1) = N(\boldsymbol{y}_1|\boldsymbol{\mu}_0, \boldsymbol{P}_0)N(\boldsymbol{x}_t|\boldsymbol{W}\boldsymbol{y}_t, \boldsymbol{\Sigma})$$

which gives

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \boldsymbol{K}_1(\boldsymbol{x}_1 - \boldsymbol{W}\boldsymbol{\mu}_0)$$
$$\boldsymbol{V}_1 = (\boldsymbol{I} - \boldsymbol{K}_1\boldsymbol{W})\boldsymbol{P}_0$$
$$\mathbf{K}_1 = \boldsymbol{P}_0\boldsymbol{W}^T(\boldsymbol{W}\boldsymbol{P}_0\boldsymbol{W}^T + \boldsymbol{\Sigma})^{-1}$$

$$c_1 = \boldsymbol{N}(\boldsymbol{x}_1|\boldsymbol{W}\boldsymbol{\mu}_0, \boldsymbol{W}\boldsymbol{P}_0\boldsymbol{W}^T + \boldsymbol{\Sigma})$$

Smoothing

Smoothing: $\gamma(\boldsymbol{y}_t) = p(\boldsymbol{y}_t|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T)$

$$\gamma(\boldsymbol{y}_t) = \hat{a}(\boldsymbol{y}_t)\hat{\beta}(\boldsymbol{y}_t) = \boldsymbol{N}(\boldsymbol{y}_t|\widehat{\boldsymbol{\mu}_t}, \widehat{\boldsymbol{V}_t})$$

We have computed $\hat{a}(\boldsymbol{y}_t)$ now let us compute $\hat{\beta}(\boldsymbol{y}_t)$ (backward step)

$$\text{HMM } c_{t+1}\hat{\beta}(\boldsymbol{z}_t) = \sum_{\boldsymbol{z}_{t+1}} \hat{\beta}(\boldsymbol{z}_{t+1})p(\boldsymbol{x}_{t+1}|\boldsymbol{z}_{t+1})p(\boldsymbol{z}_{t+1}|\boldsymbol{z}_t)$$

$$c_{t+1}\hat{\beta}(\boldsymbol{y}_t)$$

$$= \int_{\boldsymbol{y}_{t+1}} \hat{\beta}\,(\boldsymbol{y}_{t+1})\mathrm{p}(\boldsymbol{x}_{t+1}|\boldsymbol{y}_{t+1})\mathrm{p}(\boldsymbol{y}_{t+1}|\boldsymbol{y}_t)\mathbf{d}\boldsymbol{y}_{t+1}$$

Smoothing

$$\gamma(\boldsymbol{y}_t) = N(\boldsymbol{y}_t | \widehat{\boldsymbol{\mu}_t}, \widehat{\boldsymbol{V}_t})$$

After similar manipulations as in $\hat{a}(\boldsymbol{y}_t)$ we get the updates

$$\widehat{\boldsymbol{\mu}_t} = \boldsymbol{\mu}_t + \boldsymbol{J}_t(\widehat{\boldsymbol{\mu}_{t+1}} - \boldsymbol{A}\boldsymbol{\mu}_t)$$

$$\widehat{\boldsymbol{V}_t} = \boldsymbol{V}_t + \boldsymbol{J}_t(\widehat{\boldsymbol{V}_{t+1}} - \boldsymbol{P}_t)\boldsymbol{J}_t{}^T$$

$$\boldsymbol{J}_t = \boldsymbol{V}_t \boldsymbol{A}^T (\boldsymbol{P}_t)^{-1}$$

Its necessary to complete the forward step so that we have $\boldsymbol{P}_t$ and $\boldsymbol{\mu}_t$ computed

Smoothing

$$p(\boldsymbol{y}_{t-1}, \boldsymbol{y}_t | \boldsymbol{x}_1, \dots, \boldsymbol{x}_T) =$$

$$\xi(\boldsymbol{y}_{t-1}, \boldsymbol{y}_t) = (\boldsymbol{c}_t)^{-1} \hat{a}\,(\boldsymbol{y}_{t-1}) \mathrm{p}(\boldsymbol{x}_t | \boldsymbol{y}_t) \mathrm{p}(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}) \hat{\beta}\,(\boldsymbol{y}_t)$$

Which is again a Gaussian

$$\xi(\boldsymbol{y}_{t-1}, \boldsymbol{z}_t) = N(\begin{bmatrix} \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_t \end{bmatrix} | \dot{\boldsymbol{\mu}}_t, \boldsymbol{R}_t)$$

$$\dot{\boldsymbol{\mu}}_t = \begin{bmatrix} \widehat{\boldsymbol{\mu}_{t-1}} \\ \widehat{\boldsymbol{\mu}_t} \end{bmatrix} \qquad \boldsymbol{R}_t = \begin{bmatrix} \widehat{\boldsymbol{V}_{t-1}} & \boldsymbol{J}_{t-1}\widehat{\boldsymbol{V}_t} \\ (\boldsymbol{J}_{t-1}\widehat{\boldsymbol{V}_t})^T & \widehat{\boldsymbol{V}_t} \end{bmatrix}$$

E: Step

$$E[\boldsymbol{y}_t] = \int_{\boldsymbol{y}_t} N(\boldsymbol{y}_t | \widehat{\boldsymbol{\mu}_t}, \widehat{\boldsymbol{V}_t})\, d\, \boldsymbol{y}_t = \widehat{\boldsymbol{\mu}_n}$$

$$E[\boldsymbol{y}_t \boldsymbol{y}_t^{\,T}] = \widehat{\boldsymbol{V}_t} + \widehat{\boldsymbol{\mu}_t}\widehat{\boldsymbol{\mu}_t}^T$$

$$E[\boldsymbol{y}_t \boldsymbol{y}_{t-1}^{\,T}] = \widehat{\boldsymbol{V}_t}\boldsymbol{J}_{t-1}^{\,T} + \widehat{\boldsymbol{\mu}_t}\widehat{\boldsymbol{\mu}_{t-1}}^T$$

EM

Assume the sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T$

The complete likelihood is given by:

$$p(\boldsymbol{X}, \boldsymbol{Y}|\theta) = p\left(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_T, \boldsymbol{y}_{1,}\boldsymbol{y}_2, \cdots, \boldsymbol{y}_T|\theta\right)$$

$$= \prod_{t=1}^{T} p(\boldsymbol{x}_t|\boldsymbol{y}_t)\, p(\boldsymbol{y}_1) \prod_{t=2}^{T} p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1})$$

$$\underset{ln}{\Rightarrow} \ln p(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{\theta}) = \ln p(\boldsymbol{y}_1|\boldsymbol{\mu}_0, \boldsymbol{P}_0) + \sum_{t=2}^{T} \ln p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}, \mathbf{A}, \boldsymbol{\Gamma})$$

$$+ \sum_{t=1}^{T} \ln p(\boldsymbol{x}_t|\boldsymbol{y}_t, \mathbf{W}, \boldsymbol{\Sigma})$$

EM

Now we take the expectation with regards to Y|X

$$\underset{E}{\Rightarrow} \mathrm{E}[\ln p(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{\theta})]$$

$$\mathrm{E}[\ln p(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{\theta})] = E[\ln p(\boldsymbol{y}_1|\boldsymbol{\mu}_0, \boldsymbol{P}_0)] + E[\sum_{t=2}^{T} \ln p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}, \mathbf{A}, \boldsymbol{\Gamma})]$$

$$+E[\sum_{t=1}^{T} \ln p(\boldsymbol{x}_t|\boldsymbol{y}_t, \mathbf{W}, \boldsymbol{\Sigma})]$$

EM

To find $\boldsymbol{\mu}_0, \boldsymbol{P}_0$ we need the first term only

$$E[\ln p(\boldsymbol{y}_1|\boldsymbol{\mu}_0, \boldsymbol{P}_0)] = -\frac{1}{2}\ln|\boldsymbol{P}_\mathbf{0}| - E\left[\frac{1}{2}(\boldsymbol{y}_1-\boldsymbol{\mu}_0)^{\boldsymbol{T}}\boldsymbol{P}_\mathbf{0}^{-\mathbf{1}}(\boldsymbol{y}_1-\boldsymbol{\mu}_0)\right]$$

Taking the derivative of the above and making equal to zero we get

$$\boldsymbol{\mu}_0{}^{new} = \mathrm{E}[\boldsymbol{y}_1]$$

$$\mathbf{P}_0{}^{new} = \mathrm{E}[\boldsymbol{y}_1\boldsymbol{y}_1{}^T] - \mathrm{E}[\boldsymbol{y}_1]\mathrm{E}[\boldsymbol{y}_1{}^T]$$

EM

To find A, Γ we need the second term only

$$E[\sum_{t=2}^{N} \ln p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}, \mathbf{A}, \boldsymbol{\Gamma})] = -\frac{N-1}{2} \ln |\boldsymbol{\Gamma}|$$

$$-E\left[\frac{1}{2}\sum_{\boldsymbol{t=2}}^{\boldsymbol{N}} (\boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{y}_{t-1})^{\boldsymbol{T}}\, \boldsymbol{\Gamma}^{-1} (\boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{y}_{t-1})\right]$$

$$\boldsymbol{A}^{new} = \left(\sum_{t=2}^{T} \mathrm{E}\,[\boldsymbol{y}_t \boldsymbol{y}_{t-1}{}^T]\right)\left(\sum_{t=2}^{T} \mathrm{E}\,[\boldsymbol{y}_{t-1} \boldsymbol{y}_{t-1}{}^T]\right)^{-1}$$

$$\boldsymbol{\Gamma}^{new} = \frac{1}{\mathrm{N}-1}\sum_{t=2}^{T}\{E[\boldsymbol{y}_t \boldsymbol{y}_t{}^T] - \boldsymbol{A}^{new} E[\boldsymbol{y}_{t-1} \boldsymbol{y}_t{}^T]$$
$$- E[\boldsymbol{y}_t \boldsymbol{y}_{t-1}{}^T](\boldsymbol{A}^{new})^T E[\boldsymbol{y}_{t-1} \boldsymbol{y}_{t-1}{}^T](\boldsymbol{A}^{new})^T\}$$

EM

To find C, W we need the third term only

$$E\left[\sum_{t=1}^{T} \ln p(\boldsymbol{x}_t|\boldsymbol{y}_t, \mathbf{W}, \boldsymbol{\Sigma})\right]$$

$$= -\frac{N}{2}\ln|\Sigma| - E\left[\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{W}\boldsymbol{y}_t)^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x}_t - \boldsymbol{W}\boldsymbol{y}_t)\right]$$

EM

Taking the derivative and forcing to zero we get

$$\boldsymbol{W}^{new} = \left(\sum_{t=1}^{T} \boldsymbol{x}_t \mathrm{E}\left[\boldsymbol{y}_t{}^T\right]\right)\left(\sum_{t=1}^{T} \mathrm{E}\left[\boldsymbol{y}_t \boldsymbol{y}_t{}^T\right]\right)^{-1}$$

$$\boldsymbol{\Sigma}^{new} = \frac{1}{\mathrm{T}} \sum_{t=1}^{T} \{\boldsymbol{x}_t \boldsymbol{x}_t{}^T - \boldsymbol{W}^{new} \mathrm{E}[\boldsymbol{y}_t] \boldsymbol{x}_t{}^T - \boldsymbol{x}_t \mathrm{E}[\boldsymbol{y}_t{}^T] \boldsymbol{W}^{new} + (\boldsymbol{W}^{new})^T \mathrm{E}[\boldsymbol{y}_t \boldsymbol{y}_t{}^T] \boldsymbol{W}^{new}\}$$