1. a) Let $y_t' = y_t \wedge y_{t-1} = y_t, y_{t-1}$

Then, $Pr(y_t' \mid y'_{t-1})$

$= Pr(y_t' = (a,b) \mid y'_{t-1} = (c,d))$

$= Pr(y_t = a, y_{t-1} = b \mid y_{t-1} = c, y_{t-2} = d)$

$= \begin{cases} 0 & b \neq c \\ P(y_t = a \mid y_{t-1} = c, y_{t-2} = d) = P(y_t \mid y_{t-1}, y_{t-2}) & b = c \end{cases}$

$Pr(x_t \mid y_t')$

$= Pr(x_t = a \mid y_t' = (b,c))$

$= Pr(x_t = a \mid y_t = b, y_{t-1} = c)$

$= Pr(x_t = a \mid y_t = b)$ (x_t only depends on y_t)

$= Pr(x_t \mid y_t)$

$\Rightarrow$ We have a HMM parameterized by $Pr(y_t' \mid y'_{t-1})$ & $Pr(x_t \mid y_t')$.

b) Yes, because the reparameterized system has a non-zero parameter only for each non-zero parameter in the initial system.

More concretely, if there are N distinct output classes,

System	$Pr(y_t \mid y_{t-1}, y_{t-2})$ $Pr(x_t \mid y_t)$	$Pr(y_t' \mid y'_{t-1})$ $Pr(x_t \mid y_t')$
Initial State	$N^2 - 1$ ($Pr(y_0, y_1)$)	$N^2 - 1$ ($Pr(y_1') = Pr(y_0, y_1)$)
Transition	$N^2(N-1)$ ($Pr(y_t \mid y_{t-1}, y_{t-2})$)	$N^2(N-1)$ (Only non-zero parameters are $Pr(y_t \mid y_{t-1}, y_{t-2})$).
Emission	$2N$ ($Pr(x_t \mid y_t) \leftarrow$ 2 Gaussian Params)	$2N$ ($Pr(x_t \mid y_t') = Pr(x_t \mid y_t)$)

$\Rightarrow$ Both systems have $(N^2 - 1) + N^2(N-1) + 2N = N^3 + 2N - 1$ parameters.

$\Rightarrow$ No increase in parameters.

② $Pr(x_{1\ldots T}) = \sum_{y_1,\ldots,y_T} Pr(x_{1\ldots T} \mid y_{1\ldots T})\, Pr(y_{1\ldots T})$

$$= \sum_{y_1,\ldots,y_T} \left[\prod_{i=1}^{T} Pr(x_i \mid y_i)\; Pr(y_1) \prod_{i=2}^{T} P(y_i \mid y_{i-1}) \right]$$

~~* For 2 classes~~ $\cancel{c_1, c_2}$, ~~and two outputs~~ $\cancel{v_1, v_2}$

$\cancel{\# c_1^{start}}$

$\cancel{\# c_i}$

$\cancel{\#(c_i, c_j)}$

$\cancel{\#(v_i, \theta_j^{c})}$

$$\cancel{= \sum_{y_1,\ldots,y_T} \left[(\pi_{c_1})^{\# c_1^{start}} (1-\pi_{c_1})^{\# c_2^{start}} (\theta_{c_1|c_1})^{\#(c_1,c_1)} (1-\theta_{c_1|c_1})^{\#(c_2,c_1)} (\theta_{c_2|c_2})^{\#(c_1,c_2)} (1-\theta_{c_1|c_2})^{\#(c_2,c_2)} \right.}$$

$$= \sum_{y_1,\ldots,y_T} \left[\prod_{i=1}^{T} \phi_{x_i|y_i} \prod_{i=2}^{T} \theta_{y_i|y_{i-1}}\, \pi_{y_1} \right]$$

$$\Rightarrow \pi^*, \theta^*, \phi^* = \underset{\pi,\theta,\phi}{\arg\max} \sum_{y_1,\ldots,y_T} \left[\pi_{y_1} \prod_{i=1}^{T} \phi_{x_i|y_i} \prod_{i=2}^{T} \theta_{y_i|y_{i-1}} \right]$$

$$= \underset{\pi,\theta,\phi}{\arg\max} \sum_{y_1=c_1}^{c_N} \sum_{y_2=c_2}^{c_N} \cdots \sum_{y_T=c_1}^{c_N} \left[\pi_{y_1} \prod_{i=1}^{T} \phi_{x_i|y_i} \prod_{i=2}^{T} \theta_{y_i|y_{i-1}} \right]$$

If $\{c_1, \ldots, c_N\}$ are the classes.