1) a) From Appendix we have:
$$\begin{cases} p(z) = \mathcal{N}(z \mid \mu, \Sigma) \\ p(x \mid z) = \mathcal{N}(x \mid Az + b, S) \\ p(x) = \mathcal{N}(x \mid A\mu + b, A\Sigma A^T + S) \\ p(z \mid x) = \mathcal{N}(z \mid C(A^T S^{-1}(x-b) + \Sigma^{-1}\mu), C) \end{cases}$$

$$\left.\begin{array}{l} z \sim \mathcal{N}(0,1) \\ p(z) = \mathcal{N}(z \mid \mu, \Sigma) \end{array}\right\} \Rightarrow \boxed{\mu = 0, \ \Sigma = 1}$$

$$\left.\begin{array}{l} x \mid z \sim \mathcal{N}(zu, \sigma^2 I) \\ p(x \mid z) = \mathcal{N}(x \mid Az + b, S) \end{array}\right\} \Rightarrow \begin{array}{l} Az + b = zu \\ S = \sigma^2 I \end{array} \Rightarrow \boxed{\begin{array}{l} b = 0 \\ A = u \\ S = \sigma^2 I \end{array}}$$

Therefore we have:
$$p(z) = \mathcal{N}(0,1)$$
$$p(x \mid z) = \mathcal{N}(x \mid uz, \sigma^2 I)$$
$$p(x) = \mathcal{N}(x \mid 0, \sigma^2 I + uu^T)$$
$$p(z \mid x) = \mathcal{N}(z \mid C(u^T(\sigma^2 I)^{-1} x), C) = \mathcal{N}(z \mid C(\tfrac{1}{\sigma^2} u^T x), C)$$
$$C = (1 + u^T(\sigma^2 I)^{-1} u)^{-1} = (1 + \tfrac{1}{\sigma^2} u^T u)^{-1}$$

$$m = E[z \mid x] = C(\tfrac{1}{\sigma^2} u^T x) = \frac{\frac{1}{\sigma^2} u^T x}{1 + \frac{1}{\sigma^2} u^T u}$$

$$s = E[z^2 \mid x] = C + E[z \mid x]^2 = C + m^2$$

b) $$u_{new} \leftarrow \arg\max_{u} \frac{1}{N} \sum_{i=1}^{N} E_{q(z^{(i)})} \left[ \log p(z^{(i)}, x^{(i)}) \right]$$

By expanding out $\log p(z^{(i)}, x^{(i)})$ :

$$u_{new} \leftarrow \arg\max_{u} \frac{1}{N} \sum E_{q(z^{(i)})} \left[ \log p(x^{(i)} | z^{(i)}) + \log p(z^{(i)}) \right]$$

By distributing the expectation :

$$u_{new} \leftarrow \arg\max_{u} \frac{1}{N} \sum_{i=1}^{N} E_{q(z^{(i)})} \left[ \log p(x^{(i)} | z^{(i)}) \right] + E_{q(z^{(i)})} \left[ \log p(z^{(i)}) \right]$$

Gaussian formula with zero mean and unit variance:

$$N(z, \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2} \cdot \frac{(z-\mu)^2}{\sigma^2}\right) \Rightarrow N(z, 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right)$$

By substituting $N(z, 0, 1)$ with $p(z^{(i)})$ :

$$u_{new} \leftarrow \arg\max_{u} \frac{1}{N} \sum_{i=1}^{N} \left( E_{q(z^{(i)})} \left[ \log p(x^{(i)} | z^{(i)}) \right] + E_{q(z^{(i)})} \left[ \log \left( (2\pi)^{-1/2} \cdot \exp\left(-\frac{1}{2} (z^{(i)})^2\right) \right] \right) \right)$$

$$u_{new} \leftarrow \arg\max_{u} \frac{1}{N} \sum_{i=1}^{N} \left( \underbrace{E_{q(z^{(i)})} \left[ \log p(x^{(i)} | z^{(i)}) \right]}_{A} + E_{q(z^{(i)})} \left[ -\frac{1}{2} \left[ \log(2\pi) + (z^{(i)})^2 \right] \right] \right)$$

continue ->

Because we have $X|Z \sim N(uz, \sigma^2 I)$

$$E_{q(z^{(i)})}\left[\log \frac{1}{2\pi^{D/2}\sqrt{|\sigma^2 I|}} \exp\left[(x^{(i)} - z^{(i)}u)^T(\sigma^2 I)^{-1}(x^{(i)} - z^{(i)}u)\right]\right]$$

$$= E_{q(z^{(i)})}\left[-\frac{D}{2}\log 2\pi - \frac{D}{2}\log \sigma^2 + \frac{1}{\sigma^2}(x^{(i)} - z^{(i)}u)^T(x^{(i)} - z^{(i)}u)\right]$$

$$= E_{q(z^{(i)})}\Big[\underbrace{-\frac{D}{2}\log 2\pi - \frac{D}{2}\log \sigma^2 + \frac{1}{\sigma^2} x^{(i)T}x^{(i)}}_{B^{(i)}} - x^{(i)T}z^{(i)}u - u^T z^{(i)T}x^{(i)}$$

$$+ u^T z^{(i)T} z^{(i)} u\Big]$$

$$= E_{q(z^{(i)})}\left[B^{(i)} - 2z^{(i)}x^{(i)T}u + z^{(i)2}u^Tu\right]$$

$$= E_{q(z^{(i)})}[B^{(i)}] - 2x^{(i)T}u\,E_{q(z^{(i)})}[z^{(i)}] + u^Tu\,E_{q(z^{(i)})}[z^{(i)2}]$$

$$A = E_{q(z^{(i)})}[B^{(i)}] - 2m^{(i)}x^{(i)T}u + s^{(i)}u^Tu$$

There fore by substituting $A$ in $u_{new}$:

$$u_{new} \leftarrow \arg\max_u \frac{1}{N}\sum_{i=1}^{N}\left(E_{q(z^{(i)})}[B^{(i)}] - 2m^{(i)}x^{(i)T}u + s^{(i)}u^Tu\right) +$$

$$E_{q(z^{(i)})}\left[-\frac{1}{2}\left[\log(2\pi) + z^{(i)2}\right]\right]\Big)$$

$$\frac{\partial}{\partial u}\sum_{i=1}^{N}\left(E_{q(z^{(i)})}[B^{(i)}] - 2m^{(i)}x^{(i)T}u + s^{(i)}u^Tu\right) + E_{q(z^{(i)})}\left[-\frac{1}{2}\log(2\pi) + z^{(i)2}\right] = 0$$

$$\Rightarrow -2\sum_{i=1}^{N} m^{(i)}x^{(i)} + 2u\sum_{i=1}^{N} s^{(i)} = 0 \Rightarrow u_{new} = \frac{\sum_{i=1}^{N} m^{(i)}x^{(i)}}{\sum_{i=1}^{N} s^{(i)}}$$

**2.** We want to show that:

$$\| T^{\pi} Q_1(s,a) - T^{\pi} Q_2(s,a) \|_{\infty} \leq \gamma \, \| Q_1(s',a') - Q_2(s',a') \|_{\infty}$$

We start with the left part:

$$\max_{s,a} | T^{\pi} Q_1(s,a) - T^{\pi} Q_2(s,a) | = \max_{s,a} \Big| \Big[ r(s,a) + \gamma \sum_{s'} P(s'|a,s) \sum_{a'} \pi(a'|s') \, Q_1(s',a') \Big]$$

$$- \Big[ r(s,a) + \gamma \sum_{s'} P(s'|a,s) \sum_{a'} \pi(a'|s') \, Q_2(s',a') \Big] \Big|$$

By factoring:

$$= \Big| \gamma \sum_{s'} P(s'|a,s) \Big[ \sum_{a'} \pi(a'|s') \, Q_1(s',a') - \sum_{a'} \pi(a'|s') \, Q_2(s',a') \Big] \Big|$$

$$= \Big| \gamma \underbrace{\sum_{s'} P(s'|a,s)}_{\text{This is equal to 1 because its the sum of all probabilities}} \sum_{a'} \pi(a'|s') \big[ Q_1(s',a') - Q_2(s',a') \big] \Big|$$

$$\leq \gamma \max_{s'} \Big| \sum_{a'} \pi(a'|s') \big[ Q_1(s',a') - Q_2(s',a') \big] \Big|$$

$$\leq \gamma \max_{s'} \Big| \max_{a'} \big[ Q_1(s',a') - Q_2(s',a') \big] \Big|$$

We apply the absolute value to the term inside max, so it is greater than the previous term.

$$\leq \gamma \max_{s',a'} | Q_1(s',a') - Q_2(s',a') |$$

3.

a)

$$r(S,A) = \begin{cases} 1 & \text{if } S = S_1 \\ 2 & \text{if } S = S_2 \end{cases}$$

| | stay | switch |
|---|---|---|
| $S_1$ | | ✓ |
| $S_2$ | ✓ | |

$\gamma = 0.9$

$Q(S_1, \text{stay}) = 1 + \gamma Q(S_1, \text{switch})$

$Q(S_1, \text{switch}) = 1 + \gamma Q(S_2, \text{switch})$

$Q(S_2, \text{stay}) = 2 + \gamma Q(S_2, \text{stay})$

$Q(S_2, \text{switch}) = 2 + \gamma Q(S_1, \text{switch})$

by solving the 4 equations we get:

$$\begin{cases} Q(S_1, \text{Stay}) = 18.1 \\ Q(S_1, \text{switch}) = 19 \\ Q(S_2, \text{stay}) = 20 \\ Q(S_2, \text{switch}) = 19.1 \end{cases}$$

| | stay | switch |
|---|---|---|
| $S_1$ | 18.1 | 19 |
| $S_2$ | 20 | 19.1 |

b) We want to define function Q in a way that is equilibrium and also stays at S1:

$Q(S_1, \text{stay}) = 1 + \gamma Q(S_1, \text{stay})$

$0.1\, Q(S_1, \text{stay}) = 1 \Rightarrow Q(S_1, \text{stay}) = 10$

If we set the value of $Q(S_1, \text{switch})$ to a value less than 10, the agent will stay at S1 and this would be a suboptimal policy.

For example here is a table showing the Q function:

| | stay | switch |
|---|---|---|
| $S_1$ | 10 | 0 |
| $S_2$ | 0 | 0 |