test

Extra Credit

a. We will show that $(\theta^T T^{-1} \mu)^T = \mu^T T^{-1} \theta$

First, we know for matrices A, B, and C where ABC is defined, $(ABC)^T = C^T B^T A^T$.

Proof $(ABC)^T = C^T B^T A^T$:

We know this because we know for $i \times k$ matrix A and $k \times j$ matrix B, AB exists and $(AB)_{ij} = \sum_k A_{ik} B_{kj}$.

We also know $(AB)^T_{ij} = (AB)_{ji} = \sum_k A_{jk} B_{ki} = \sum_k A^T_{kj} B^T_{ik}$
$= \sum_k B^T_{ik} A^T_{kj} = (B^T A^T)_{ij}$. So $(AB)^T = B^T A^T$.

Similarly, for A, B, and $j \times l$ matrix C, $(ABC)_{il} = \sum_j (AB)_{ij} (C)_{jl}$.
And $(ABC)^T_{il} = (ABC)_{li} = \sum_j (AB)_{lj} (C)_{ji} = \sum_j (AB)^T_{jl} (C)^T_{ij}$
$= \sum_j C^T_{ij} (AB)^T_{jl} = (C^T (AB)^T)_{il} = (C^T B^T A^T)_{il}$. So $(ABC)^T = C^T B^T A^T$.

So we know $(\theta^T T^{-1} \mu)^T = (\mu)^T (T^{-1})^T (\theta^T)^T$. We can simplify further.
First, since for any matrix X, $(X^T)^T = X$, $(\theta^T)^T = \theta$.
Next, we know the inverse of a symmetric matrix is also symmetric, so $(T^{-1})^T = T^{-1}$!
We can derive this result: $TT^{-1} = I$, and $I = I^T$, so $(TT^{-1}) = (TT^{-1})^T$,
or $(TT^{-1}) = (T^{-1})^T T^T$.

Also, $T^{-1}T = TT^{-1} = I$, so $T^{-1}T = (T^{-1})^T T^T$.
And $T = T^T$, since T is symmetric, so $T^{-1}T = (T^{-1})^T T$
$\hookrightarrow T^{-1} T T^{-1} = (T^{-1})^T T T^{-1}$
$\hookrightarrow T^{-1} I = (T^{-1})^T I$
$T^{-1} = (T^{-1})^T$.

So since T is symmetric, $(T^{-1})^T = T^{-1}$. So $\boxed{(\theta^T T^{-1} \mu)^T = \mu^T T^{-1} \theta}$ ✓

Also, $\theta^T T^{-1} \mu$ is a scalar (because of the dimensions being 1×1), and the transpose of a scalar is the same scalar, so $\theta^T T^{-1} \mu = \mu^T T^{-1} \theta$.

b $p(\theta) = (2\pi)^{-p/2} (T^{-1})^{-1/2} e^{-\frac{1}{2}(\theta-\mu)^T (T^{-1})(\theta-\mu)}$

$$\propto e^{-\frac{1}{2}(\theta^T T^{-1}\theta + \theta^T T^{-1}\mu - \underbrace{\mu^T T^{-1}\theta} + \mu^T T^{-1}\mu)}$$

Based on a, $\mu^T T^{-1}\theta = \theta^T T^{-1}\mu$.

$$\propto e^{-\frac{1}{2}(\theta^T T^{-1}\theta \underbrace{- \theta^T T^{-1}\mu - \theta^T T^{-1}\mu}_{\text{Combine like terms}} + \mu^T T^{-1}\mu)}$$

$$\propto e^{-\frac{1}{2}(\theta^T T^{-1}\theta - 2\theta^T T^{-1}\mu + \underbrace{\mu^T T^{-1}\mu}_{\text{Can be canceled out, no } \theta \text{ term}})}$$

$$\propto \boxed{e^{-\frac{1}{2}(\theta^T T^{-1}\theta - 2\theta^T T^{-1}\mu)}}$$ ✓

c $p(\theta \mid \Sigma, y_{1:n}) \propto p(y_{1:n} \mid \theta, \Sigma) \cdot p(\theta)$

We can use b: $p(\theta) \propto e^{-\frac{1}{2}(\theta^T T^{-1}\theta - 2\theta^T T^{-1}\mu)}$

$$p(y_{1:n} \mid \theta, \Sigma) = \prod_{i=1}^{n} (2\pi)^{-p/2} (\Sigma)^{-1/2} e^{-\frac{1}{2}(y_i-\theta)^T \Sigma^{-1}(y_i-\theta)}$$

$$\propto e^{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta)}$$

$$\propto e^{-\frac{1}{2}\sum_{i=1}^{n}(y_i^T\Sigma^{-1}y_i \underbrace{- y_i^T\Sigma^{-1}\theta}_{\text{Equal}} \underbrace{- \theta^T\Sigma^{-1}y_i} + \theta^T\Sigma^{-1}\theta)}$$

$$\propto e^{-\frac{1}{2}\sum_{i=1}^{n}(-2\theta^T\Sigma^{-1}y_i + \theta^T\Sigma^{-1}\theta)}$$

$$\propto e^{-\frac{1}{2}(-2\theta^T\Sigma^{-1}n\bar{y} + n\theta^T\Sigma^{-1}\theta)}$$

$$p(\theta \mid \Sigma, y_{1:n}) \propto e^{-\frac{1}{2}(-2\theta^T\Sigma^{-1}n\bar{y} + n\theta^T\Sigma^{-1}\theta)} e^{-\frac{1}{2}(\theta^T T^{-1}\theta - 2\theta^T T^{-1}\mu)}$$

$$\propto e^{-\frac{1}{2}(-2\theta^T(\Sigma^{-1}n\bar{y} + T^{-1}\mu) + \theta^T(n\Sigma^{-1}+T^{-1})\theta)}$$

$$\propto e^{-\frac{1}{2}\theta^T(n\Sigma^{-1}+T^{-1})\theta + \theta^T(n\Sigma^{-1}\bar{y} + T^{-1}\mu)}$$

$$\propto MVN((n\Sigma^{-1}+T^{-1})^{-1}(n\Sigma^{-1}\bar{y}+T^{-1}\mu), (n\Sigma^{-1}+T^{-1})^{-1})$$ ✓

So we have showed $p(\theta, \Sigma, y) \sim MVN(\mu^*(\Sigma) = T^*(n\Sigma^{-1}\bar{y} + T^{-1}\mu), T^* = (n\Sigma^{-1}+T^{-1})^{-1})$ ✓

We will prove this equality. We can also use the fact that $tr(A+B) = tr(A) + tr(B)$.

d $$\underbrace{tr((\Psi \Sigma^{-1}) + \sum_{i=1}^{n} (y_i - \theta)^T \Sigma^{-1} (y_i - \theta)}_{\text{This is a scalar } a = tr(a)} = tr(\Psi \Sigma^{-1}) + tr\left(\sum_{i=1}^{n} (y_i - \theta)^T \Sigma^{-1} (y_i - \theta)\right)$$

$$= tr(\Psi \Sigma^{-1}) + \sum_{i=1}^{n} \underbrace{tr\left(((y_i - \theta)^T \Sigma^{-1} (y_i - \theta)\right)}_{tr(ABC) = tr(CAB)}$$

$$= tr(\Psi \Sigma^{-1}) + \sum_{i=1}^{n} tr((y_i - \theta)(y_i - \theta)^T \Sigma^{-1})$$

$$= tr(\Psi \Sigma^{-1}) + tr\left(\sum_{i=1}^{n} (y_i - \theta)(y_i - \theta)^T \Sigma^{-1}\right)$$

$$= tr\left(\Psi \Sigma^{-1} + \sum_{i=1}^{n} (y_i - \theta)(y_i - \theta)^T \Sigma^{-1}\right)$$

$$= tr\left(\left(\Psi + \sum_{i=1}^{n} (y_i - \theta)(y_i - \theta)^T\right) \Sigma^{-1}\right)$$

So $tr(\Psi \Sigma^{-1}) + \sum_{i=1}^{n} (y_i - \theta)^T \Sigma^{-1} (y_i - \theta) = tr\left(\left(\Psi + \sum_{i=1}^{n} (y_i - \theta)(y_i - \theta)^T\right) \Sigma^{-1}\right)$ ✓

e $p(\Sigma \mid y, \theta) \propto p(y \mid \theta, \Sigma) \cdot p(\Sigma)$

$p(y \mid \theta, \Sigma) \propto \det(\Sigma)^{-n/2} e^{-\frac{1}{2}\left(\sum_{i=1}^{n} (y_i - \theta)^T \Sigma^{-1} (y_i - \theta)\right)}$

$p(\Sigma) \propto \det(\Sigma)^{-\frac{1}{2}(v+p+1)} e^{-\frac{1}{2}(tr(\Psi \Sigma^{-1}))}$

$p(\Sigma \mid y, \theta) \propto \det(\Sigma)^{-\frac{1}{2}(v+p+1+n)} e^{-\frac{1}{2}\underbrace{(tr(\Psi \Sigma^{-1}) + \sum_{i=1}^{n} (y_i - \theta)^T \Sigma^{-1} (y_i - \theta))}_{\text{Using the answer to d}}}$

$\propto \det(\Sigma)^{-\frac{1}{2}(v+p+1+n)} e^{-\frac{1}{2}(tr((\Psi + \sum_{i=1}^{n} (y_i - \theta)(y_i - \theta)^T)\Sigma^{-1})}$

$\Sigma \mid \theta, \text{data} \sim IW\left(n+v, \left(\Psi + \sum_{i=1}^{n} (y_i - \theta)(y_i - \theta)^T\right)^{-1}\right)$ ✓