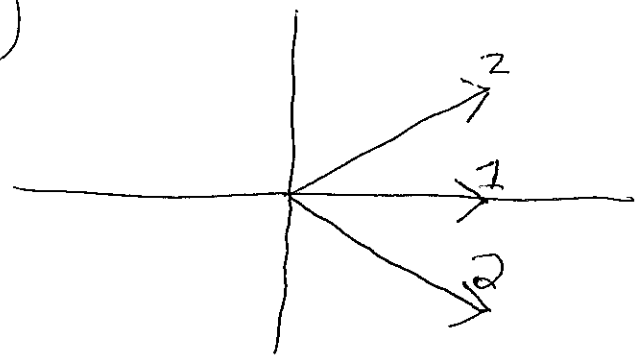
#5)



Easy to verify that the solution planes for these 3 points have a null intersection (no solution cone).

#5) $P[x_1, x_2, \ldots x_T / \mu, \Sigma] =$

$$\frac{1}{(2\pi)^{Td/2}\,|\Sigma|^{T/2}} \cdot e^{\left[-\frac{1}{2}\sum_{k=1}^{T}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu)\right]}$$

$$l(\mu,\Sigma) = \log P[\cdot]$$

$$= -\frac{Td}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{k=1}^{T}\left(x_k^T\Sigma^{-1}x_k - 2\mu^T\Sigma^{-1}x_k + \mu^T\Sigma^{-1}\mu\right)$$

$$= -\frac{Td}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{k=1}^{T}x_k^Tx_k + \mu^T\Sigma^{-1}\cdot\sum_{k=1}^{T}x_k - \frac{T}{2}\mu^T\Sigma^{-1}\mu$$

To calculate $\nabla_\mu l(\cdot)$ and $\nabla_\Sigma l(\cdot)$, we'll use the following ∇ properties (for vectors $\underline{r}, \underline{v}$ and matrix A)

1) $\nabla_{\underline{r}}(\underline{r}^T\underline{v}) = \underline{v}$
2) $\nabla_{\underline{r}}(\underline{r}^TA\underline{r}) = (A+A^T)\underline{r}$
3) $\nabla_A(\underline{r}^TA\underline{r}) = \underline{r}\,\underline{r}^T$
4) $\nabla_A(\ln|A|) = A^{-1}$

$$\nabla_\mu(l(\mu,\Sigma)) = \nabla_\mu\Big(\mu^T\underbrace{\big(\Sigma^{-1}\cdot\sum_{k=1}^{T}x_k\big)}_{\underline{v}}\Big) - \frac{T}{2}\nabla_\mu(\mu^T\Sigma^{-1}\mu)$$

From 2), this equals $(\Sigma^{-1}+(\Sigma^{-1})^T)\mu = 2\Sigma^{-1}\mu$

$\boxed{\text{from 1), this equals } \left(\Sigma^{-1}\cdot\sum_{k=1}^{T}x_k\right)}$

So: $\nabla_\mu \ell(\hat{\mu}, \hat{\Sigma}) = \hat{\Sigma}^{-1}\left(\sum_{k=1}^{T} x_k\right) - T\hat{\Sigma}^{-1}\hat{\mu} = 0$

multiply both sides by $\hat{\Sigma}$
+ divide by $T \Longrightarrow$

$$\boxed{\hat{\mu} = \frac{1}{T}\sum_{k=1}^{T} x_k}$$

$\nabla_\Sigma(\ell(\mu, \Sigma)) = ?$

To simplify this part, let $B = \Sigma^{-1}$.

Then, $\ell(\mu, B) = -\frac{Td}{2}\ln(2\pi) - \frac{T}{2}\ln|B^{-1}|$
$\quad - \frac{1}{2}\sum_{k=1}^{T}(x_k - \mu)^T B (x_k - \mu)$

$\ln|B^{-1}| = -\ln|B|$

Also, from 4), $\nabla_B \ln|B| = B^{-1} \Longrightarrow$

$\nabla_B\left(-\frac{T}{2}\ln|B^{-1}|\right) = \frac{T}{2}B^{-1}$

Next, $\nabla_B\left(-\frac{1}{2}\sum_{k=1}^{T}(x_k-\mu)^T B(x_k-\mu)\right) =$

$-\frac{1}{2}\sum_{k=1}^{T}\nabla_B\left[(x_k-\mu)^T B(x_k-\mu)\right] \overset{\text{From 3)}}{=}$

$-\frac{1}{2}\sum_{k=1}^{T}(x_k-\mu)(x_k-\mu)^T$

So: $\nabla_B(\ell(\hat{\mu}, \hat{B})) = \frac{T}{2}\hat{B}^{-1} - \frac{1}{2}\sum_{k=1}^{T}(x_k-\hat{\mu})(x_k-\hat{\mu})^T = 0$

$\hat{B}^{-1} = \hat{\Sigma} \Longrightarrow$

$$\boxed{\hat{\Sigma} = \frac{1}{T}\sum_{k=1}^{T}(x_k-\hat{\mu})(x_k-\hat{\mu})^T}$$

#7) Using the notation from class:

$$\frac{\partial}{\partial \theta_1}(l_k) = \frac{1}{\theta_2}(X_k - \theta_1)$$

$$\frac{\partial}{\partial \theta_2}(l_k) = -\frac{1}{2\theta_2} + \frac{1}{2\theta_2^2}(X_k - \theta_1)^2$$

$$\frac{\partial^2 (l_k)}{\partial \theta_1^2} = -\frac{1}{\theta_2}$$

$$\frac{\partial^2 (l_k)}{\partial \theta_2^2} = \frac{1}{2\theta_2^2} - \frac{1}{\theta_2^3}(X_k - \theta_1)^2$$

$$\frac{\partial^2 (l_k)}{\partial \theta_1 \partial \theta_2} = -\frac{(X_k - \theta_1)}{\theta_2^2}$$

$$\frac{\partial^2 (l_k)}{\partial \theta_2 \partial \theta_1} = -\frac{1}{\theta_2^2}(X_k - \theta_1)$$

Now, $$\frac{\partial^2 l(\underline{\theta})}{\partial \theta_i \partial \theta_j} = \sum_{k=1}^{T} \frac{\partial^2 (l_k)}{\partial \theta_i \partial \theta_j}, \quad \forall i, j$$

$$H = \left.\begin{pmatrix} \frac{\partial^2 l(\underline{\theta})}{\partial \theta_1^2} & \frac{\partial^2 l(\underline{\theta})}{\partial \theta_2 \partial \theta_1} \\ \frac{\partial^2 l(\underline{\theta})}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 l(\underline{\theta})}{\partial \theta_2^2} \end{pmatrix}\right|_{\underline{\theta} = \hat{\theta}_{ML}}$$

It is easy to see that, in our case,

$$H = \begin{pmatrix} -\frac{N}{\hat{\sigma}^2} & 0 \\ 0 & \underbrace{\frac{N}{2\hat{\sigma}^4} - \frac{N}{\hat{\sigma}^4}}_{-\frac{N}{2\hat{\sigma}^4}} \end{pmatrix}$$

Let $\underline{a} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$ be an arbitrary 2-vector.

Then, $\underline{a}^T H \underline{a} = -\frac{N}{\hat{\sigma}^2} a_0^2 - \frac{N}{2\hat{\sigma}^4} a_1^2 < 0$

H negative definite $\Longrightarrow$ ML solution is a maximum...