What is the relationship between

- purity criterion?
- log-likelihood criterion?

The data sets are generated following the below parameter setting:

$$\boldsymbol{y_{ki}} = \boldsymbol{X_i}(\boldsymbol{\beta_k} + \boldsymbol{b_{ki}} + \boldsymbol{\Gamma_k}(\boldsymbol{\alpha_i'x_i})) + \boldsymbol{\epsilon_{ki}} \tag{1}$$

where $\boldsymbol{\beta_k}$ and $\boldsymbol{\Gamma_k(\alpha_i'x_i)}$ present the fixed effect, $\boldsymbol{b}_{ki}$ is the random effect and $\boldsymbol{\epsilon_{ki}}$ present the random error. We can rewrite the equation (1) as

$$\boldsymbol{y_{ki}} = \begin{pmatrix} \boldsymbol{X_i} & \boldsymbol{S_i} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta_k} \\ \boldsymbol{\Gamma_k} \end{pmatrix} + \boldsymbol{X_i b}_{ki} + \boldsymbol{\epsilon_{ki}} \tag{2}$$

where

- $\boldsymbol{X}_i, \boldsymbol{S}_i$ are $n_{ti} \times 3$ matrix and $\begin{pmatrix} \mathbf{X_i} & \mathbf{S_i} \end{pmatrix}$ is then the $n_{ti} \times 6$ design matrix (suppose it is a quadratic model with intercept, slope and concavity), n_{ti} is the number of time points.

- $\boldsymbol{S_i} = \boldsymbol{X_i}(\boldsymbol{\alpha}_i'\boldsymbol{x}_i) = \boldsymbol{X_i}(\boldsymbol{\alpha_1} + \boldsymbol{\delta\alpha_{2i}})'\boldsymbol{x_i}$, $\boldsymbol{\alpha}_1$ is the 3 matrix with fixed value while $\boldsymbol{\alpha}_{2i}$ is the 3 matrix generated from some distributions.

Suppose we have n subject in each of the k group, the matrix of the outcome is

$$\boldsymbol{Y}_k = \begin{pmatrix} \boldsymbol{y_{k1}} \\ \boldsymbol{y_{k2}} \\ \dots \\ \boldsymbol{y_{kn}} \end{pmatrix}_{N \times 1}$$

, where $N = \sum_{i=1}^{n} n_{ti}$, y_{ki} has the dimension $n_{ti} \times 1$.

The fixed effect design matrix is

$$\boldsymbol{Z}_k = \begin{pmatrix} \boldsymbol{X_{k1}} & \boldsymbol{S}_{k1} \\ \boldsymbol{X_{k2}} & \boldsymbol{S}_{k2} \\ \dots & \\ \boldsymbol{X_{kn}} & \boldsymbol{S}_{kn} \end{pmatrix}_{N \times 2t} = \begin{pmatrix} \boldsymbol{X_{k1}} & \boldsymbol{w}_1 X_{k1} \\ \boldsymbol{X_{k2}} & \boldsymbol{w}_2 X_{k2} \\ \dots & \\ \boldsymbol{X_{kn}} & \boldsymbol{w}_n X_{kn} \end{pmatrix}_{N \times 2t}$$

, where $w_i = \boldsymbol{\alpha}_i'\boldsymbol{x}_i$, $t = 3$ here since the we fit a quadratic model.

The design matrix for random effect is

$$\boldsymbol{X}_k = \begin{pmatrix} \boldsymbol{X_{k1}} & 0... & 0 & \\ 0 & \boldsymbol{X_{k2}} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \boldsymbol{X_{kn}} \end{pmatrix}_{N \times (n \times t)}$$

, the matrix of predictors $\boldsymbol{x}_k = \begin{pmatrix} \boldsymbol{x_{k1}} \\ \boldsymbol{x_{k2}} \\ \dots \\ \boldsymbol{x_{kn}} \end{pmatrix}_{N \times p}$ and random effect's coefficient matrix is $\boldsymbol{b}_k = \begin{pmatrix} \boldsymbol{b_{k1}} \\ \boldsymbol{b_{k2}} \\ \dots \\ \boldsymbol{b_{kn}} \end{pmatrix}_{3t \times 1}$

The design matrix for random error is

$$\boldsymbol{\epsilon}_k = \begin{pmatrix} \boldsymbol{\epsilon_{k1}} \\ \boldsymbol{\epsilon_{k2}} \\ \dots \\ \boldsymbol{\epsilon_{kn}} \end{pmatrix}_{N \times 1}$$

Therefore, the model can be written as

$$\boldsymbol{Y}_k = \boldsymbol{X}_k \otimes \beta_k + \boldsymbol{X}_k(\boldsymbol{x}_k\boldsymbol{\alpha} \otimes \boldsymbol{\Gamma}_k) + \boldsymbol{X}_k\boldsymbol{b}_k + \boldsymbol{\epsilon}_k \tag{3}$$

The estimation of fixed effect is then To make it easy, let's assume a simple scenario, where

- We only have two treatment group, $K = 2$
- Each group have n subjects
- Each subject in the group has same time points $\boldsymbol{X}_i = \boldsymbol{X}$ (X is a $n_t \times t$ matrix)
- The subjects are assigned to group 1 and group 2 with same probability $\pi_1 = \pi_2 = 0.5$

Suppose: We have n_1 subjects, who should be in group1 and are actually assigned to group 1; n_1 subjects who should be assigned to group 2 and are actually assigned to group 2. Since we set $\pi_1 = \pi_2 = 0.5$, $n_1 = \frac{n}{2}$. That is

	True: Group 1	True: Group	Total
Assigned: Group 1	$n_1 = \frac{n}{2}$	$n - n_1 = \frac{n}{2}$	n
Assigned: Group 2	$n - n_1 = \frac{n}{2}$	$n_1 = \frac{n}{2}$	n
Total	n	n	$2n$

And also, to make it easy, let's suppose the subjects $1, 2, ..., n$ are in group 1 and subjects $n+1, ..., 2n$ are in group 2.

$$\boldsymbol{y_{ki}} = \boldsymbol{X_i}(\boldsymbol{\beta_k} + \boldsymbol{b_{ki}} + \boldsymbol{\Gamma_k}(\boldsymbol{\alpha_i'x_i})) + \boldsymbol{\epsilon_{ki}} \tag{4}$$

Since we assume the outcomes follow multivariate normal distributions the PDF for a subject in group i can be written as,

$$f(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^{n_t}|\boldsymbol{\Sigma}_k|}} \exp(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu}_k)'\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_k)) \tag{5}$$

where $\boldsymbol{\Sigma}_k = \boldsymbol{X}\boldsymbol{D_k}\boldsymbol{X}' + \sigma_k^2\boldsymbol{I}$, $\boldsymbol{\mu}_k = \boldsymbol{X}(\boldsymbol{\beta_k} + \boldsymbol{\Gamma_k}(\boldsymbol{\alpha'x}))$

Therefore, the likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta}|\boldsymbol{x}) =& \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^{n_t}|\boldsymbol{\Sigma}_1|}} \exp(-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})) \\ &\times \prod_{i=n+1}^{2n} \frac{1}{\sqrt{(2\pi)^{n_t}|\boldsymbol{\Sigma}_2|}} \exp(-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})) \end{aligned} \tag{6}$$

The log-likelihood function is

$$\begin{aligned} l(\boldsymbol{\theta}|\boldsymbol{x}) =& - n \cdot n_t \log(2\pi) - \frac{n}{2}\log(|\boldsymbol{\Sigma}_1|) - \frac{n}{2}\log(|\boldsymbol{\Sigma}_2|) \\ &- \sum_{i=1}^{n} \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i}) \\ &- \sum_{i=n+1}^{2n} \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i}) \end{aligned} \tag{7}$$

So we aim to find the $\boldsymbol{\alpha}$ that maximizes the log-likelihood function. The value of the log-likelhood function depends on the outcome values $\boldsymbol{y}_i$ and the predictors values $\boldsymbol{x}_i$.

Recall our purity function,

$$Purity(\boldsymbol{\alpha}) = A_0 + A_1\boldsymbol{\mu}_x'\boldsymbol{\alpha} + \frac{A_2}{2}\left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_x\boldsymbol{\alpha} + \boldsymbol{\alpha}'\boldsymbol{\mu}_x\boldsymbol{\mu}_x'\boldsymbol{\alpha}\right] \tag{8}$$

where

- $A_0 = -t + \frac{1}{2}tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1) + \frac{1}{2}tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2) + \frac{1}{2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)$
- $A_1 = (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})(\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)$
- $A_2 = (\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2))'(\boldsymbol{D}_1^{-1} + \boldsymbol{D}_2^{-1})((\boldsymbol{\Gamma}_1 - \boldsymbol{\Gamma}_2)$

that is, the purity function only depends on the predictors values $\boldsymbol{x}_i$.

To make them comparable, how about we calculate the expectation of the log-likelihood function and find $\boldsymbol{\alpha}$ that maximizes the expectation?

$$
\begin{aligned}
E(l(\boldsymbol{\theta})|\boldsymbol{x}) =& - n \cdot n_t \log(2\pi) - \frac{n}{2}\log(|\boldsymbol{\Sigma}_1|) - \frac{n}{2}\log(|\boldsymbol{\Sigma}_2|) \\
& - n_1 E_1\big[\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})\big] - (n-n_1)E_1\big[\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})\big] \\
& - (n-n_1)E_2\big[\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})\big] - n_1 E_2\big[\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})\big] \\
=& - n \cdot n_t \log(2\pi) - \frac{n}{2}\log(|\boldsymbol{\Sigma}_1|) - \frac{n}{2}\log(|\boldsymbol{\Sigma}_2|) \\
& - \frac{n}{4}E_1\big[(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})\big] - \frac{n}{4}E_1\big[(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})\big] \\
& - \frac{n}{4}E_2\big[(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})\big] - \frac{n}{4}E_2\big[(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})\big]
\end{aligned}
\tag{9}
$$

And

$$
\begin{aligned}
E_1[(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})] =& E_1[tr((\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i}))] \\
=& E_1[tr\big(\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\big)] \\
=& tr\big(E_1[\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})']\big) \\
=& tr\big(\Sigma_1^{-1}E_1[(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})']\big) \\
=& tr\big(\Sigma_1^{-1}\Sigma_1\big) = tr(\boldsymbol{I}_{n_t}) = n_t \ (n_t \text{ is the number of time points})
\end{aligned}
\tag{10}
$$

Similarly

$$
E_2\big[(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})\big] - \frac{n}{4}E_2\big[(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})\big] = n_t \tag{11}
$$

On the other hand

$$
\begin{aligned}
& E_1[(\boldsymbol{y}_i - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(y_i - \boldsymbol{\mu}_{2i})] \\
=& E_1[(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i} + \boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i} + \boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})] \\
=& E_1[(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i}) + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i}) \\
& + (\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i}) + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})] \\
=& E_1[(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})] + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}E_1(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i}) + \\
& E_1(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\Sigma_2^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i}) + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i}) \\
=& E_1[(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})] + 0 + 0 + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i}) \\
=& E_1[tr(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i}))] + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i}) \\
=& E_1[tr(\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i}))] + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i}) \\
=& tr(E_1[\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})]) + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i}) \\
=& tr(\boldsymbol{\Sigma}_2^{-1}E_1[(y_i - \boldsymbol{\mu}_{1i})'(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})]) + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i}) \\
=& tr(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})
\end{aligned}
\tag{12}
$$

Similarly

$$
E_2[(\boldsymbol{y}_i - \boldsymbol{\mu}_{1i})'\boldsymbol{\Sigma}_1^{-1}(y_i - \boldsymbol{\mu}_{1i})] = tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2) + (\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i}) \tag{13}
$$

Therefore

$$
\begin{aligned}
E(l(\boldsymbol{\theta})|\boldsymbol{x}) \propto & \frac{n}{2}\log(|\boldsymbol{\Sigma}_1|) - \frac{n}{2}\log(|\boldsymbol{\Sigma}_2|) - \frac{n}{4}tr(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1) \\
& - \frac{n}{4}tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2) - \frac{n}{4}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})'(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{2i})
\end{aligned}
\tag{14}
$$

And

$$
\begin{aligned}
&(\boldsymbol{\mu}_{1i}-\boldsymbol{\mu}_{2i})'(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{\mu}_{1i}-\boldsymbol{\mu}_{2i})\\
&=\big(\boldsymbol{X}\boldsymbol{\beta}_1-\boldsymbol{X}\boldsymbol{\beta}_2+(\boldsymbol{X}\boldsymbol{\Gamma}_1-\boldsymbol{X}\boldsymbol{\Gamma}_2)\boldsymbol{\alpha}'\boldsymbol{x}_i\big)'(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})\big(\boldsymbol{X}\boldsymbol{\beta}_1-\boldsymbol{X}\boldsymbol{\beta}_2+(\boldsymbol{X}\boldsymbol{\Gamma}_1-\boldsymbol{X}\boldsymbol{\Gamma}_2)\boldsymbol{\alpha}'\boldsymbol{x}_i\big)\\
&=(\boldsymbol{X}\boldsymbol{\beta}_1-\boldsymbol{X}\boldsymbol{\beta}_2)'(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{X}\boldsymbol{\beta}_1-\boldsymbol{X}\boldsymbol{\beta}_2)\\
&+2\big[(\boldsymbol{X}\boldsymbol{\beta}_1-\boldsymbol{X}\boldsymbol{\beta}_2)'(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{X}\boldsymbol{\Gamma}_1-\boldsymbol{X}\boldsymbol{\Gamma}_2)\boldsymbol{x}_i'\boldsymbol{\alpha}\\
&+\boldsymbol{\alpha}'\boldsymbol{x}_i\boldsymbol{x}_i'\boldsymbol{\alpha}\big((\boldsymbol{X}\boldsymbol{\Gamma}_1-\boldsymbol{X}\boldsymbol{\Gamma}_2)\big)'(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})\big((\boldsymbol{X}\boldsymbol{\Gamma}_1-\boldsymbol{X}\boldsymbol{\Gamma}_2)\big)
\end{aligned}
\tag{15}
$$

Therefore, the expectation of the log-likelihood is

$$
\begin{aligned}
&E(l(\boldsymbol{\theta}))\propto-\frac{n}{2}\log(|\boldsymbol{\Sigma}_1|)-\frac{n}{2}\log(|\boldsymbol{\Sigma}_2|)-\frac{n}{4}tr(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1)-\frac{n}{4}tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_2)\\
&-\frac{n}{4}\big\{(\boldsymbol{X}\boldsymbol{\beta}_1-\boldsymbol{X}\boldsymbol{\beta}_2)'(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{X}\boldsymbol{\beta}_1-\boldsymbol{X}\boldsymbol{\beta}_2)\\
&+2\big[(\boldsymbol{X}\boldsymbol{\beta}_1-\boldsymbol{X}\boldsymbol{\beta}_2)'(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{X}\boldsymbol{\Gamma}_1-\boldsymbol{X}\boldsymbol{\Gamma}_2)\boldsymbol{\mu}_x'\boldsymbol{\alpha}\\
&+\boldsymbol{\alpha}'(\boldsymbol{\mu_x}\boldsymbol{\mu_x'}+\boldsymbol{\Sigma_x})\boldsymbol{\alpha}\big((\boldsymbol{X}\boldsymbol{\Gamma}_1-\boldsymbol{X}\boldsymbol{\Gamma}_2)\big)'(\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_2^{-1})\big((\boldsymbol{X}\boldsymbol{\Gamma}_1-\boldsymbol{X}\boldsymbol{\Gamma}_2)\big)\big\}
\end{aligned}
\tag{16}
$$

Purity

$$
\begin{aligned}
Purity(\boldsymbol{\alpha})\propto&\frac{1}{2}tr(\boldsymbol{D}_2^{-1}\boldsymbol{D}_1)+\frac{1}{2}tr(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2)+\frac{1}{2}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)\\
&+(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})(\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\boldsymbol{\mu}_x'\boldsymbol{\alpha}\\
&+\frac{1}{2}(\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)'(\boldsymbol{D}_1^{-1}+\boldsymbol{D}_2^{-1})\big((\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2)\big[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_x\boldsymbol{\alpha}+\boldsymbol{\alpha}'\boldsymbol{\mu}_x\boldsymbol{\mu}_x'\boldsymbol{\alpha}\big]
\end{aligned}
$$