1 Derivations

To simplify the question, we focus on the derivation of covariance structure to one cluster only. In actual implementation, we need to deal with a mixer of multiple clusters. Assume we have a data set $\mathbf{X}$ of size n in $\mathbf{R}^d$. Each observation $\mathbf{x}^{(i)}$ is a d by 1 vector drawn from the distribution $\mathbf{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. log likelihood can be calculated by

$$
\begin{aligned}
l(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X}) &= log \prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} exp(-\frac{1}{2}(\mathbf{x}^{(i)} - \mu)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}^{(i)} - \mu)) \\
&= -\frac{n}{2} log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}^{(i)} - \mu)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}^{(i)} - \mu) + constant \\
&= -\frac{n}{2} log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n} \mathbf{t}^{(i)^T} \boldsymbol{\Sigma}^{-1} \mathbf{t}^{(i)} + constant
\end{aligned}
$$

For simplicity, we let $\mathbf{t}^{(i)} = \mathbf{x}^{(i)} - \mu$.

1.1 Constant Case

With constant structure covariance, we can restructure $\boldsymbol{\Sigma}$ as $s\mathbf{I}_{d\times d}$, where s is a positive value.

$$
\begin{aligned}
\frac{\partial}{\partial s}[-\frac{n}{2}log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)^T}\boldsymbol{\Sigma}^{-1}\mathbf{t}^{(i)}] &= \frac{\partial}{\partial s}[-\frac{n}{2}log|\mathbf{sI}| - \frac{1}{2s}\sum_{i=1}^{n}\mathbf{t}^{(i)^T}\mathbf{t}^{(i)}] \\
&= \frac{\partial}{\partial s}[-\frac{nd}{2}log(s) - \frac{1}{2s}\sum_{i=1}^{n}\mathbf{t}^{(i)^T}\mathbf{t}^{(i)}] \\
&= 0
\end{aligned}
$$

Solve for s we get

$$
\begin{aligned}
0 &= \frac{\partial}{\partial s}[-\frac{nd}{2}log(s) - \frac{1}{2s}\sum_{i=1}^{n}\mathbf{t}^{(i)^T}\mathbf{t}^{(i)}] \\
\frac{nd}{s} &= \frac{1}{s^2}\sum_{i=1}^{n}\mathbf{t}^{(i)^T}\mathbf{t}^{(i)} \\
s &= \frac{\sum_{i=1}^{n}\mathbf{t}^{(i)^T}\mathbf{t}^{(i)})}{nd}
\end{aligned}
$$

1.2 Diagonal Case

When the covariance structure is diagonal, we can view $\boldsymbol{\Sigma}$ as a d-dimensional vector $\{s_1,\ s_2\ ...\ s_d\}$. Let $\mathbf{M}$ be the inverse of $\boldsymbol{\Sigma}$, so we can denote $\mathbf{M}$ and its elements as $\{m_1,\ m_2\ ...\ m_d\}$. The derivative of the kth ($k \leq d$) element of $\mathbf{\Sigma}$ can be derived as follows

$$\frac{\partial}{\partial m_k}[\frac{n}{2}log(m_1 m_2 ... m_d) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{d} m_j {t_j^{(i)}}^2] = \frac{n}{2m_k} - \frac{1}{2}\sum_{i=1}^{n} {t_k^{(i)}}^2$$

Here t_k is the kth element of the vector $\mathbf{t}$. When the above equation is set as 0, we get

$$\frac{n}{2m_k} - \frac{1}{2}\sum_{i=1}^{n} {t_k^{(i)}}^2 = 0$$

$$s_k = \frac{\sum_{i=1}^{n} {t_k^{(i)}}^2}{n}$$

1.3 Mixed Term Structure:

Let $\mathbf{\Sigma^{-1}} = \lambda\mathbf{M} + (1-\lambda)s\mathbf{I}$ where $\mathbf{M}$ and s are restricted by constrain $\log|\mathbf{M}| = d\log s$. The reason that we chose to equate $\mathbf{\Sigma^{-1}}$ to $\lambda\mathbf{M} + (1-\lambda)s\mathbf{I}$ instead of using $\mathbf{\Sigma}$ will be discussed in Appendix 1. We attempt to use Lagrange multiplier to solve for $\mathbf{M}$ and s. The object function that we want to maximize will be the log likelihood function of $\mathbf{M}$ and s

$$f(\mathbf{M}, s) = l(\boldsymbol{\mu}, \mathbf{\Sigma}|\mathbf{X}) = \frac{n}{2}log|(\lambda\mathbf{M} + (\mathbf{1}-\lambda)\mathbf{sI})| - \frac{1}{2}\sum_{i=1}^{n} {\mathbf{t}^{(i)}}^T(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})\mathbf{t}^{(i)}$$

under constraint such that

$$g(\mathbf{M}, s) = d\log s - \log|\mathbf{M}| = 0$$

The Lagrangian problem is to maximize $f(\mathbf{M}, s)$ such that

$$\begin{cases} \nabla f(\mathbf{M}, s) = \alpha\nabla g(\mathbf{M}, s) \\ d\log s - \log|\mathbf{M}| = 0 \end{cases}$$

Since

$$\frac{\partial \log|\mathbf{M}|}{\partial \mathbf{M}} = \mathbf{M}^{-1}$$

$$\frac{\partial}{\partial s} d\log s = \frac{d}{s}$$

The gradient of the constrain function is

$$\begin{pmatrix} \mathbf{M}_{1,1}^{-1} \\ \mathbf{M}_{1,2}^{-1} \\ \vdots \\ \mathbf{M}_{d,d}^{-1} \\ \frac{d}{s} \end{pmatrix} \tag{1}$$

Here $\mathbf{M}_{i,j}^{-1}$ stands for the i,jth component of the M inverse matrix. The gradient of the object function can be solved by chain rule. First we can calculate,

$$\begin{aligned}\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X})}{\partial \boldsymbol{\Sigma}^{-1}} &= \frac{n}{2}\boldsymbol{\Sigma} - \frac{1}{2}\sum_{i=1}^{n} \mathbf{t}^{(i)}\mathbf{t}^{(i)^T} \\ &= \frac{n}{2}(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})^{-1} - \frac{1}{2}\sum_{i=1}^{n} \mathbf{t}^{(i)}\mathbf{t}^{(i)^T}\end{aligned}$$

To continue evaluate the gradient we can stretch the $\boldsymbol{\Sigma}^{-1}$ and $\mathbf{M}$ matrix into d^2 long vectors, so that we can circumvent the calculation of Jacobian matrix.

$$\begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \cdots & \Sigma_{1,d} \\ \Sigma_{2,1} & \Sigma_{2,2} & \cdots & \Sigma_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{d,1} & \Sigma_{d,2} & \cdots & \Sigma_{d,d} \end{pmatrix} \xRightarrow{\text{stretch}} (\Sigma_{1,1}, \Sigma_{1,2}, \cdots, \Sigma_{1,d}, \Sigma_{2,1}, \cdots, \Sigma_{d,d})$$

$$\begin{pmatrix} M_{1,1} & M_{1,2} & \cdots & M_{1,d} \\ M_{2,1} & M_{2,2} & \cdots & M_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ M_{d,1} & M_{d,2} & \cdots & M_{d,d} \end{pmatrix} \xRightarrow{\text{stretch}} (M_{1,1}, M_{1,2}, \cdots, M_{1,d}, M_{2,1}, \cdots, M_{d,d})$$

The derivative of $\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \mathbf{M}}$ will therefore become a d^2 by d^2 matrix

$$\begin{pmatrix} \frac{\partial \Sigma_{1,1}^{-1}}{\partial M_{1,1}} & \frac{\partial \Sigma_{1,2}^{-1}}{\partial M_{1,1}} & \cdots & \frac{\partial \Sigma_{d,d}^{-1}}{\partial M_{1,1}} \\ \frac{\partial \Sigma_{1,1}^{-1}}{\partial M_{1,2}} & \frac{\partial \Sigma_{1,2}^{-1}}{\partial M_{1,2}} & \cdots & \frac{\partial \Sigma_{d,d}^{-1}}{\partial M_{1,2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Sigma_{1,1}^{-1}}{\partial M_{d,d}} & \frac{\partial \Sigma_{1,2}^{-1}}{\partial M_{d,d}} & \cdots & \frac{\partial \Sigma_{d,d}^{-1}}{\partial M_{d,d}} \end{pmatrix} = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{pmatrix}_{d^2 \times d^2}$$

According to chain rule,

$$\begin{aligned}\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X})}{\partial \mathbf{M}} &= \frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X})}{\partial \boldsymbol{\Sigma}^{-1}} \frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \mathbf{M}} \\ &= \lambda[\frac{n}{2}(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})^{-1} - \frac{1}{2}\sum_{i=1}^{n} \mathbf{t}^{(i)}\mathbf{t}^{(i)^T}]\end{aligned}$$

The same trick can be apply to the calculation of $\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X})}{\partial s}$. By calculating with the stretched $\boldsymbol{\Sigma}^{-1}$ vector, we will get

$$\begin{aligned}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial s} &= (\nabla_s \boldsymbol{\Sigma}_{1,1}^{-1}, \nabla_s \boldsymbol{\Sigma}_{1,2}^{-1}, \cdots, \nabla_s \boldsymbol{\Sigma}_{1,d}^{-1}, \nabla_s \boldsymbol{\Sigma}_{2,1}^{-1}, \nabla_s \boldsymbol{\Sigma}_{2,2}^{-1}, \cdots, \nabla_s \boldsymbol{\Sigma}_{2,d}^{-1}, \cdots, \nabla_s \boldsymbol{\Sigma}_{d,d}^{-1}) \\ &= (\overbrace{\underbrace{1-\lambda, 0, \cdots, 0}_{\text{d}}, \underbrace{0, 1-\lambda, \cdots, 0}_{\text{d}}, \cdots, 1-\lambda}^{d^2})\end{aligned}$$

According to chain rule

$$\frac{\partial l(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X})}{\partial s} = (1-\lambda)\,\mathrm{Tr}(\frac{n}{2}\boldsymbol{\Sigma} - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)}{\mathbf{t}^{(i)}}^{T})$$

Therefore, $\nabla f(\mathbf{M}, s)$ is

$$\nabla f(\mathbf{M}, s) = \begin{pmatrix} \lambda[\frac{n}{2}(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})^{-1} - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)}{\mathbf{t}^{(i)}}^{T}] \\ (1-\lambda)\,\mathrm{Tr}[\frac{n}{2}(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})^{-1} - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)}{\mathbf{t}^{(i)}}^{T}] \end{pmatrix}$$

Under the setting of Lagrange multiplier,

$$\lambda[\frac{n}{2}\boldsymbol{(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})^{-1}} - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)}{\mathbf{t}^{(i)}}^{T}] = \alpha\mathbf{M}^{-1}$$

$$(1-\lambda)\,\mathrm{Tr}[\frac{n}{2}(\boldsymbol{\lambda\mathbf{M} + (1-\lambda)s\mathbf{I}})^{-1} - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)}{\mathbf{t}^{(i)}}^{T}] = \alpha\frac{d}{s}$$

$$d\log s - \log|\mathbf{M}| = 0$$

Replace s with $|\mathbf{M}|^{\frac{1}{d}}$

$$\lambda[\boldsymbol{\frac{n}{2}(\lambda\mathbf{M} + (1-\lambda)|\mathbf{M}|^{\frac{1}{d}}\mathbf{I})^{-1} - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)}{\mathbf{t}^{(i)}}^{T}}] = \alpha\mathbf{M}^{-1} \tag{2}$$

$$(1-\lambda)\,\mathrm{Tr}[\frac{n}{2}(\boldsymbol{\lambda\mathbf{M} + (1-\lambda)|\mathbf{M}|^{\frac{1}{d}}\mathbf{I}})^{-1} - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)}{\mathbf{t}^{(i)}}^{T}] = \alpha d|\mathbf{M}|^{-\frac{1}{d}} \tag{3}$$

Assume $\sum_{i=1}^{n}\mathbf{t}^{(i)}{\mathbf{t}^{(i)}}^{T}$ and $\mathbf{M}$ share the same eigenvectors

$$\sum_{i=1}^{n}\mathbf{t}^{(i)}{\mathbf{t}^{(i)}}^{T} = n\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathbf{-1}}$$

$$\mathbf{M} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{\mathbf{-1}}$$

We can rewrite equation (2) as

$$\lambda[\frac{n}{2}(\lambda\mathbf{M} + (1-\lambda)|\mathbf{M}|^{\frac{1}{d}}\mathbf{I})^{-1} - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)}{\mathbf{t}^{(i)}}^{T}] = \alpha\mathbf{M}^{-1}$$

$$\frac{\lambda}{2}[n(\lambda\mathbf{Q}\mathbf{D}\mathbf{Q}^{-1} + (1-\lambda)|\mathbf{D}|^{\frac{1}{d}}\mathbf{Q}\mathbf{Q}^{-1})^{-1} - n\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}] = \alpha\mathbf{Q}\mathbf{D}^{-1}\mathbf{Q}^{-1}$$

$$\frac{n\lambda}{2}[(\mathbf{Q}(\lambda\mathbf{D})\mathbf{Q}^{-1} + \mathbf{Q}((1-\lambda)|\mathbf{D}|^{\frac{1}{d}}\mathbf{I})\mathbf{Q}^{-1})^{-1} - \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}] = \alpha\mathbf{Q}\mathbf{D}^{-1}\mathbf{Q}^{-1}$$

$$\frac{n\lambda}{2}[(\mathbf{Q}(\lambda\mathbf{D})\mathbf{Q}^{-1} + \mathbf{Q}((1-\lambda)|\mathbf{D}|^{\frac{1}{d}}\mathbf{I})\mathbf{Q}^{-1})^{-1} - \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}] = \alpha\mathbf{Q}\mathbf{D}^{-1}\mathbf{Q}^{-1}$$

$$\frac{n\lambda}{2}[\mathbf{Q}(\lambda\mathbf{D} + (1-\lambda)|\mathbf{D}|^{\frac{1}{d}}\mathbf{I})^{-1}\mathbf{Q}^{-1} - \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}] = \alpha\mathbf{Q}\mathbf{D}^{-1}\mathbf{Q}^{-1}$$

$$\lambda[(\lambda\mathbf{D} + (1-\lambda)|\mathbf{D}|^{\frac{1}{d}}\mathbf{I})^{-1} - \boldsymbol{\Lambda}] = \frac{2\alpha}{n}\mathbf{D}^{-1}$$

Here let D_i denotes the ith diagonal element of $\mathbf{D}$

$$\lambda[(\lambda D_i + (1-\lambda)\prod_{i=1}^{d} D_i^{\frac{1}{d}})^{-1} - \Lambda_i] = \frac{2\alpha}{n}\frac{1}{D_i}$$

We can rewrite equation (3) as

$$(1-\lambda)\,\mathrm{Tr}[\frac{n}{2}(\boldsymbol{\lambda}\mathbf{M} + \mathbf{(1-\lambda)}|\mathbf{M}|^{\frac{1}{d}}\mathbf{I})^{-1} - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)}\mathbf{t}^{(i)T}] = \alpha d|\mathbf{M}|^{-\frac{1}{d}}$$

$$(1-\lambda)\,\mathrm{Tr}[n\mathbf{Q}(\lambda\mathbf{D} + (1-\lambda)|\mathbf{D}|^{\frac{1}{d}}\mathbf{I})^{-1}\mathbf{Q}^{-1} - n\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}] = \alpha d|\mathbf{D}|^{-\frac{1}{d}}$$

$$n(1-\lambda)\,\mathrm{Tr}[(\lambda\mathbf{D} + (1-\lambda)|\mathbf{D}|^{\frac{1}{d}}\mathbf{I})^{-1}] - n(1-\lambda)\,\mathrm{Tr}(\boldsymbol{\Lambda}) = \alpha d|\mathbf{D}|^{-\frac{1}{d}}$$

Here let D_i denotes the ith diagonal element of matrix $\mathbf{D}$

$$(1-\lambda)\sum_{i=1}^{d}[(\lambda D_i + (1-\lambda)\prod_{i=1}^{d} D_i^{\frac{1}{d}})^{-1} - \Lambda_i] = \frac{\alpha d}{n}\prod_{i=1}^{d} D_i^{\frac{-1}{d}}$$

Theoretically, we can solve for α and D_i based on the system of equation

$$\begin{cases} (1-\lambda)\sum_{i=1}^{d}[(\lambda D_i + (1-\lambda)\prod_{i=1}^{d} D_i^{\frac{1}{d}})^{-1} - \Lambda_i] = \frac{\alpha d}{n}\prod_{i=1}^{d} D_i^{\frac{-1}{d}} \\ \lambda[(\lambda D_1 + (1-\lambda)\prod_{j=1}^{d} D_j^{\frac{1}{d}})^{-1} - \Lambda_1] = \frac{2\alpha}{n}\frac{1}{D_1} \\ \lambda[(\lambda D_2 + (1-\lambda)\prod_{j=1}^{d} D_j^{\frac{1}{d}})^{-1} - \Lambda_2] = \frac{2\alpha}{n}\frac{1}{D_2} \\ \vdots \\ \lambda[(\lambda D_d + (1-\lambda)\prod_{j=1}^{d} D_j^{\frac{1}{d}})^{-1} - \Lambda_d] = \frac{2\alpha}{n}\frac{1}{D_d} \end{cases}$$

I attempted to solve for the system of equations numerically by Newton-Gauss algorithm. Codes are documented in a python demonstration file.

2 Appendix

$\boldsymbol{\Sigma} = \lambda\mathbf{M} + (1-\lambda)s\mathbf{I}$ is a valid format with easy interpretation; however it is difficult to solve. Let $\boldsymbol{\Sigma} = \lambda\mathbf{M} + (1-\lambda)s\mathbf{I}$ where $\mathbf{M}$ and s are restricted by constrain $\log|\mathbf{M}| = d\log s$. We attempt to use Lagrange multiplier to solve for $\mathbf{M}$ and s. The object function that we want to maximize will be the log likelihood function of $\mathbf{M}$ and s under constraints.

$$f(\mathbf{M}, s) = l(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X}) = -\frac{n}{2}log|(\lambda\mathbf{M} + (\mathbf{1}-\lambda)\mathbf{sI})| - \frac{1}{2}\sum_{i=1}^{n}\mathbf{t}^{(i)T}(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})^{-1}\mathbf{t}^{(i)}$$

The derivative of the target function can be solved by chain rule,

$$\frac{\partial l(\boldsymbol{\mu}, \mathbf{\Sigma}|\mathbf{X})}{\partial \mathbf{\Sigma}} = -\frac{n}{2}\mathbf{\Sigma}^{-1} - \frac{1}{2}\frac{\partial \sum_{i=1}^{n} \mathbf{t}^{(i)^T}\mathbf{\Sigma}^{-1}\mathbf{t}^{(i)}}{\partial \mathbf{\Sigma}} \tag{4}$$

Since $\mathbf{\Sigma}$ is symmetric,

$$\begin{aligned}\frac{d}{dx_{ij}}\mathbf{A^T}\mathbf{\Sigma}^{-1}\mathbf{A} &= \mathbf{A^T}\mathbf{\Sigma}^{-1}\mathbf{e_i}\mathbf{e_j}^{\mathbf{T}}\mathbf{\Sigma}^{-1}\mathbf{A} \\ &= \mathbf{e_i}^{\mathbf{T}}\mathbf{\Sigma}^{-1}\mathbf{A}\mathbf{A^T}\mathbf{\Sigma}^{-1}\mathbf{e_j}\end{aligned}$$

Here $\mathbf{e_i}$ is the ith column of $\mathbf{I}$. Therefore,

$$\frac{d}{d\Sigma}\mathbf{A^T}\mathbf{\Sigma}^{-1}\mathbf{A} = \mathbf{\Sigma}^{-1}\mathbf{A}\mathbf{A^T}\mathbf{\Sigma}^{-1}$$

Continue equation (1),

$$\begin{aligned}\frac{\partial l(\boldsymbol{\mu}, \mathbf{\Sigma}|\mathbf{X})}{\partial \mathbf{\Sigma}} &= -\frac{n}{2}\mathbf{\Sigma}^{-1} - \frac{1}{2}\frac{\partial \sum_{i=1}^{n} \mathbf{t}^{(i)^T}\mathbf{\Sigma}^{-1}\mathbf{t}^{(i)}}{\partial \mathbf{\Sigma}} \\ &= -\frac{n}{2}\mathbf{\Sigma}^{-1} - \frac{1}{2}\sum_{i=1}^{n}\mathbf{\Sigma}^{-1}\mathbf{t}^{(i)}\mathbf{t}^{(i)^T}\mathbf{\Sigma}^{-1}\end{aligned}$$

$\frac{\partial \mathbf{\Sigma}}{\partial \mathbf{M}}$ generates a $p \times p \times p \times p$ tensor. By the same trick we used in the previous section, we can expand the tensor into a long vector, and then take the derivative element-wise. The final derivative will be

$$\frac{n\lambda}{2}(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})^{-1} + \sum_{i=1}^{d}\frac{\lambda}{2}(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})\mathbf{t_i}\mathbf{t_i}^T(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})$$

Similarly, the same trick can be applied to $\frac{\partial f}{\partial s}$.

$$\frac{\partial f}{\partial s} = \frac{n(1-\lambda)}{2}\,\mathrm{Tr}((\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})^{-1}) + \sum_{i=1}^{d}\frac{(1-\lambda)}{2}\mathbf{t_i}^T(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})^{-1}(\lambda\mathbf{M} + (1-\lambda)s\mathbf{I})^{-1}\mathbf{t_i}$$

However, these equation are hard to simplify compared to the other formatting.