1 预备知识

$$\|\boldsymbol{A}-\boldsymbol{B}\|_F^2=\operatorname{tr}\left[(\boldsymbol{A}-\boldsymbol{B})^{\mathrm{T}}(\boldsymbol{A}-\boldsymbol{B})\right]$$

$$\|\boldsymbol{A}\|_F^2=\operatorname{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$$

$$\operatorname{tr}(\boldsymbol{A}^{\mathrm{T}})=\operatorname{tr}(\boldsymbol{A})$$

$$\frac{\partial\operatorname{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})}{\partial\boldsymbol{A}}=\boldsymbol{X}$$

$$\frac{\partial\operatorname{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{X})}{\partial\boldsymbol{D}}=\frac{\partial\operatorname{tr}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{D})}{\partial\boldsymbol{D}}=\frac{\partial\operatorname{tr}\left[(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{X})^{\mathrm{T}}\right]}{\partial\boldsymbol{D}}$$

对满足矩阵乘法条件的任意 $\boldsymbol{A}_{\mathrm{m\times n}}$ 和 $\boldsymbol{B}_{\mathrm{n\times m}}$ 有 $\operatorname{tr}(\boldsymbol{AB})=\operatorname{tr}(\boldsymbol{BA})$

求导:$\nabla_{\boldsymbol{X}}\|\boldsymbol{X}\|_F^2=2\boldsymbol{X}$。若 $\boldsymbol{Y}=\boldsymbol{AX}+\boldsymbol{B}$, 其中 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 为常数矩阵, 则 $\nabla_{\boldsymbol{X}}f(\boldsymbol{AX}+\boldsymbol{B})=\boldsymbol{A}^{\mathrm{T}}\nabla_{\boldsymbol{Y}}f(\boldsymbol{Y})$

2 公式推导过程

原式为:

$$\min_{\boldsymbol{C}_i}\left\|\tilde{\boldsymbol{L}}-\sum_{i=1}^{K}\boldsymbol{C}_i^{\mathrm{T}}\tilde{\boldsymbol{X}}_i\right\|_F^2+\lambda\sum_{i=1}^{K}\|\boldsymbol{C}_i\|_F \tag{1}$$

可转换为:

$$\begin{aligned}&\min\nolimits_{\boldsymbol{C}_i}\left\|\tilde{\boldsymbol{L}}-\sum_{i=1}^{K}\boldsymbol{D}_i^{\mathrm{T}}\tilde{\boldsymbol{X}}_i\right\|_F^2+\lambda\sum_{i=1}^{K}\|\boldsymbol{C}_i\|_F\\&s.t.\boldsymbol{D}_i=\boldsymbol{C}_i\end{aligned} \tag{2}$$

写出其朗格朗日增广矩阵:

$$\begin{aligned}\Gamma\left(\boldsymbol{C}_i,\boldsymbol{D}_i,\boldsymbol{P}_i,\mu\right)=&\sum_{i=1}^{K}\left\|\tilde{\boldsymbol{L}}-\boldsymbol{D}_i^{\mathrm{T}}\tilde{\boldsymbol{X}}_i\right\|_F^2+\lambda\sum_{i=1}^{K}\|\boldsymbol{C}_i\|_F\\&+\sum_{i=1}^{K}\operatorname{tr}\left[\boldsymbol{P}_i^{\mathrm{T}}\left(\boldsymbol{C}_i-\boldsymbol{D}_i\right)\right]+\sum_{i=1}^{K}\frac{\mu}{2}\|\boldsymbol{C}_i-\boldsymbol{D}_i\|_F^2\end{aligned} \tag{3}$$

其求解过程如下所示:

- 保持其他参数不变，更新 $\boldsymbol{D}$

  Equ. (3) 可转化为:

$$\min_{\boldsymbol{D}_i}\sum_{i=1}^{K}\left\|\tilde{\boldsymbol{L}}-\boldsymbol{D}_i^{\mathrm{T}}\tilde{\boldsymbol{X}}_i\right\|_F^2+\sum_{i=1}^{K}\operatorname{tr}\left[\boldsymbol{P}_i^{\mathrm{T}}(\boldsymbol{C}_i-\boldsymbol{D}_i)\right]+\sum_{i=1}^{K}\frac{\mu}{2}\|\boldsymbol{C}_i-\boldsymbol{D}_i\|_F^2$$

由于 $\boldsymbol{C}_i$ 和 $\boldsymbol{P}_i$ 是常数矩阵，所以可以转化 (化简时注意 F 范数和 F 范数平方的区别)，为：

$$\min_{\boldsymbol{D}} \left\|\tilde{\boldsymbol{L}} - \boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}}\right\|_F^2 + \operatorname{tr}\left[\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{C}-\boldsymbol{D})\right] + \frac{\mu}{2}\left\|\boldsymbol{C}-\boldsymbol{D}\right\|_F^2$$

因为

$$\left\|\tilde{\boldsymbol{L}} - \boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}}\right\|_F^2 = \operatorname{tr}\left[(\tilde{\boldsymbol{L}} - \boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}})^{\mathrm{T}}(\tilde{\boldsymbol{L}} - \boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}})\right] + \operatorname{tr}\left[\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{C}-\boldsymbol{D})\right] + \frac{\mu}{2}\left\|\boldsymbol{C}-\boldsymbol{D}\right\|_F^2$$

其中，第一部分：

$$\begin{aligned}
① &= \operatorname{tr}\left[(\tilde{\boldsymbol{L}} - \boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}})^{\mathrm{T}}(\tilde{\boldsymbol{L}} - \boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}})\right] \\
&= \operatorname{tr}\left[(\tilde{\boldsymbol{L}}^{\mathrm{T}} - \tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D})(\tilde{\boldsymbol{L}} - \boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}})\right] \\
&= \operatorname{tr}\left[\tilde{\boldsymbol{L}}^{\mathrm{T}}\tilde{\boldsymbol{L}} - \tilde{\boldsymbol{L}}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}} - \tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D}\tilde{\boldsymbol{L}} + \tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}}\right] \\
&= \operatorname{tr}(\tilde{\boldsymbol{L}}^{\mathrm{T}}\tilde{\boldsymbol{L}}) - \operatorname{tr}(\tilde{\boldsymbol{L}}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}}) - \operatorname{tr}(\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D}\tilde{\boldsymbol{L}}) + \operatorname{tr}(\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}}) \\
&= \operatorname{tr}(\tilde{\boldsymbol{L}}^{\mathrm{T}}\tilde{\boldsymbol{L}}) - 2\operatorname{tr}(\tilde{\boldsymbol{L}}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}}) + \operatorname{tr}(\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D}) \\
&= \operatorname{tr}(\tilde{\boldsymbol{L}}^{\mathrm{T}}\tilde{\boldsymbol{L}}) - 2\operatorname{tr}(\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{L}}^{\mathrm{T}}) + \left\|\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D}\right\|_F^2
\end{aligned} \tag{4}$$

对第一部分求导：

$$\begin{aligned}
&\frac{\partial(\operatorname{tr}(\tilde{\boldsymbol{L}}^{\mathrm{T}}\tilde{\boldsymbol{L}}) - 2\operatorname{tr}(\boldsymbol{D}^{\mathrm{T}}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{L}}^{\mathrm{T}}) + \left\|\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D}\right\|_F^2)}{\partial\boldsymbol{D}} \\
&= -2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{L}}^{\mathrm{T}} + 2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D}
\end{aligned} \tag{5}$$

对第二部分求导，由于 $\boldsymbol{C}_i$ 和 $\boldsymbol{P}_i$ 是常数矩阵：

$$② = \frac{\partial\operatorname{tr}\left[\boldsymbol{P}^{\mathrm{T}}(\boldsymbol{C}-\boldsymbol{D})\right]}{\partial\boldsymbol{D}} = -\frac{\partial\operatorname{tr}\left[\boldsymbol{D}^{\mathrm{T}}\boldsymbol{P}\right]}{\partial\boldsymbol{D}} = -\boldsymbol{P}$$

对第三部分求导：

$$③ = -\frac{\mu}{2}2(\boldsymbol{C}-\boldsymbol{D}) = \mu(\boldsymbol{D}-\boldsymbol{C})$$

综上：

$$-2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{L}}^{\mathrm{T}} + 2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{D} - \boldsymbol{P} + \mu(\boldsymbol{D}-\boldsymbol{C}) = \boldsymbol{0}$$

所以：

$$(\mu\boldsymbol{I} + 2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}})\boldsymbol{D} = \boldsymbol{P} + 2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{L}}^{\mathrm{T}} + \mu\boldsymbol{C}$$

整理可有：

$$\boldsymbol{D} = (\mu\boldsymbol{I} + 2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}})^{-1}(2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{L}}^{\mathrm{T}} + \boldsymbol{P} + \mu\boldsymbol{C})$$

可得：

$$\boldsymbol{D} = (\boldsymbol{I} + 2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^{\mathrm{T}}/\mu)^{-1}\left[\boldsymbol{C} + \frac{1}{\mu}(\boldsymbol{P} + 2\tilde{\boldsymbol{X}}\tilde{\boldsymbol{L}}^{\mathrm{T}})\right]$$

- 保持其他参数不变，更新 $\boldsymbol{C}$

$$\min_{\boldsymbol{C}_i}\lambda\sum_{i=1}^{K}\left\|\boldsymbol{C}_i\right\|_F + \sum_{i=1}^{K}\operatorname{tr}\left[\boldsymbol{P}_i^{\mathrm{T}}(\boldsymbol{C}_i-\boldsymbol{D}_i)\right] + \sum_{i=1}^{K}\frac{\mu}{2}\left\|\boldsymbol{C}_i-\boldsymbol{D}_i\right\|_F^2 \tag{6}$$

凑以便求 $\boldsymbol{C}$, 某些步骤求 min 略去常数项:

$$\begin{aligned}
&\min_{\boldsymbol{C}_i} \lambda \sum_{i=1}^{K} \|\boldsymbol{C}_i\|_F + \sum_{i=1}^{K} \operatorname{tr}\left[\boldsymbol{P}_i^{\mathrm{T}}(\boldsymbol{C}_i - \boldsymbol{D}_i)\right] + \sum_{i=1}^{K} \frac{\mu}{2} \|\boldsymbol{C}_i - \boldsymbol{D}_i\|_F^2 \\
&= \min_{\boldsymbol{C}_i} \mu \frac{\lambda}{\mu} \sum_{i=1}^{K} \|\boldsymbol{C}_i\|_F + \frac{\mu}{2} \sum_{i=1}^{K} \operatorname{tr}\left[(\frac{2\boldsymbol{P}_i^{\mathrm{T}}}{\mu} + \boldsymbol{C}_i^{\mathrm{T}} - \boldsymbol{D}_i^{\mathrm{T}})(\boldsymbol{C}_i - \boldsymbol{D}_i)\right] \\
&= \mu \min_{\boldsymbol{C}_i} \left\{ \frac{\lambda}{\mu} \sum_{i=1}^{K} \|\boldsymbol{C}_i\|_F + \frac{1}{2} \sum_{i=1}^{K} \operatorname{tr}\left[\left((\boldsymbol{C}_i^{\mathrm{T}} - \boldsymbol{D}_i^{\mathrm{T}} + \frac{\boldsymbol{P}_i^{\mathrm{T}}}{\mu}) + \frac{\boldsymbol{P}_i^{\mathrm{T}}}{\mu}\right)\left((\boldsymbol{C}_i - \boldsymbol{D}_i + \frac{\boldsymbol{P}_i}{\mu}) - \frac{\boldsymbol{P}_i}{\mu}\right)\right] \right\} \\
&= \mu \min_{\boldsymbol{C}_i} \left\{ \frac{\lambda}{\mu} \sum_{i=1}^{K} \|\boldsymbol{C}_i\|_F + \frac{1}{2} \sum_{i=1}^{K} \left\|\boldsymbol{C}_i - (\boldsymbol{D}_i - \frac{\boldsymbol{P}_i}{\mu})\right\|_F^2 + \frac{1}{2} \sum_{i=1}^{K} \operatorname{tr}\left[-\frac{\boldsymbol{C}_i^{\mathrm{T}} \boldsymbol{P}}{\mu} + \frac{\boldsymbol{P}_i^{\mathrm{T}} \boldsymbol{C}_i}{\mu}\right] \right\} \\
&= \mu \min_{\boldsymbol{C}_i} \left\{ \frac{\lambda}{\mu} \sum_{i=1}^{K} \|\boldsymbol{C}_i\|_F + \frac{1}{2} \sum_{i=1}^{K} \left\|\boldsymbol{C}_i - (\boldsymbol{D}_i - \frac{\boldsymbol{P}_i}{\mu})\right\|_F^2 \right\}
\end{aligned} \tag{7}$$

对每块有:

$$\min_{\boldsymbol{C}_i} \left\{ \frac{\lambda}{\mu} \|\boldsymbol{C}_i\|_F + \frac{1}{2} \left\|\boldsymbol{C}_i - (\boldsymbol{D}_i - \frac{\boldsymbol{P}_i}{\mu})\right\|_F^2 \right\}$$

. 我们知道，形如下式

$$\min_{\boldsymbol{C}_i} \frac{\lambda}{\mu} \|\boldsymbol{C}_i\|_F + \frac{1}{2} \left\|\boldsymbol{C}_i - (\boldsymbol{Z}_i + \frac{1}{\mu}\boldsymbol{T}_i)\right\|_F^2$$

有解:

$$\boldsymbol{C}_i = \begin{cases} \dfrac{\left\|\boldsymbol{Z}_i + \dfrac{\boldsymbol{T}_i}{\mu}\right\|_F - \dfrac{\lambda}{\mu}}{\left\|\boldsymbol{Z}_i + \dfrac{\boldsymbol{T}_i}{\mu}\right\|_F} (\boldsymbol{Z}_i + \dfrac{\boldsymbol{T}_i}{\mu}), & \dfrac{\lambda}{\mu} < \left\|\boldsymbol{Z}_i + \dfrac{\boldsymbol{T}_i}{\mu}\right\|_F \\ \boldsymbol{0}, & \dfrac{\lambda}{\mu} \geq \left\|\boldsymbol{Z}_i + \dfrac{\boldsymbol{T}_i}{\mu}\right\|_F \end{cases} \tag{8}$$

所以，式子如下:

$$\min_{\boldsymbol{C}_i} \frac{\lambda}{\mu} \|\boldsymbol{C}_i\|_F + \frac{1}{2} \left\|\boldsymbol{C}_i - (\boldsymbol{D}_i - \frac{\boldsymbol{P}_i}{\mu})\right\|_F^2$$

有解:

$$\boldsymbol{C}_i = \begin{cases} \dfrac{\left\|\boldsymbol{D}_i - \dfrac{\boldsymbol{P}_i}{\mu}\right\|_F - \dfrac{\lambda}{\mu}}{\left\|\boldsymbol{D}_i - \dfrac{\boldsymbol{P}_i}{\mu}\right\|_F} (\boldsymbol{D}_i - \dfrac{\boldsymbol{P}_i}{\mu}), & \dfrac{\lambda}{\mu} < \left\|\boldsymbol{D}_i - \dfrac{\boldsymbol{P}_i}{\mu}\right\|_F \\ \boldsymbol{0}, & \dfrac{\lambda}{\mu} \geq \left\|\boldsymbol{D}_i - \dfrac{\boldsymbol{P}_i}{\mu}\right\|_F \end{cases} \tag{9}$$

令 $\mathrm{d}_i = \left\| \boldsymbol{D}_i - \frac{\boldsymbol{P}_i}{\mu} \right\|_F$，将 d_i 排序，使得 $\mathrm{d}_{i_1} \geq \mathrm{d}_{i_2} \geq \cdots \geq \mathrm{d}_{i_K}$
令 $\lambda = \mu \mathrm{d}_{i_{\kappa+1}}$, 更新 $\boldsymbol{C}^{\mathrm{T}} = [\boldsymbol{C}_1^{\mathrm{T}}, \cdots, \boldsymbol{C}_K^{\mathrm{T}}]$

- 保持其他参数不变，更新 $\boldsymbol{P}$ 和 μ

$$\boldsymbol{P} = \boldsymbol{P} + \mu(\boldsymbol{D} - \boldsymbol{C})$$

$$\mu = \min(\rho\mu, \mu_{max})$$