1. SUPPLEMENTARY MATERIAL

1.1. Group Sparsemax

原始优化问题为：

$$\max_{p\in\Delta^d} \frac{1}{2}(y-\sum_{i=1}^{n} p_i)^2+\lambda\sum_{i=1}^{n}||p_i||_2, \tag{1}$$

其中$p = \{p_1, p_2, p_3, ..., p_n\}$是生成的概率分布,$p_i = \{\theta_i^1, \theta_i^2, \theta_i^3, ..., \theta_i^m\}$是第$i$个组，$\theta_i^j$ 是第i个概率分组中的第j个概率元素,$\Delta^d = \{P \in R^d | 1^T P \le 1, P \ge 0\}$是$d$维的概率单纯形。

将其转换为Group Lasso形式:

$$\begin{aligned}
\max \quad & \frac{1}{2}(y-\sum_{i=1}^{n} H_i p_i)^2+\lambda\sum_{i=1}^{n}||p_i||_2 \\
=&\frac{1}{2}(y-\sum_{i=1}^{n} H_i p_i)^T(y-\sum_{i=1}^{n} H_i p_i)+\lambda\sum_{i=1}^{n}||p_i||_2 \\
=&\frac{1}{2}(y^T-\sum_{i=1}^{n} p_i^T H_i^T)(y-\sum_{i=1}^{n} H_i p_i)+\lambda\sum_{i=1}^{n}||p_i||_2 \\
=&\frac{1}{2}(y^T y-y^T\sum_{i=1}^{n} H_i p_i-\sum_{i=1}^{n} p_i^T H_i^T y+ \\
&\sum_{i=1}^{n}\sum_{j=1}^{n} p_i^T H_i^T H_j p_j)+\lambda\sum_{i=1}^{n}||p_i||_2 \\
=&\frac{1}{2}(y^T y-2y^T\sum_{i=1}^{n} p_i+\sum_{i=1}^{n} p_i^T p_i)+\lambda\sum_{i=1}^{n}||p_i||_2 \\
s.t \quad & 1^T p = 1, \\
& p \ge 0,
\end{aligned} \tag{2}$$

其中$H_i = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$.

Eq. 2的Larangian形式如下：

$$\begin{aligned}
\mathcal{L}(y,\mu,\tau) =&\frac{1}{2}(y^T y-2y^T\sum_{i=1}^{n} p_i+\sum_{i=1}^{n} p_i^T p_i)+ \\
&\lambda\sum_{i=1}^{n}||p_i||_2-\mu p+\tau(1^T p-1).
\end{aligned} \tag{3}$$

同时最优解$\{p^*, \mu^*, \tau^*\}$需要满足下列的KKT条件：

1) 当$p_i > 0$时，

$$p^* \ge 0, \mu^* \ge 0, \mu^* p = 0, 1^T p^* = 1 \tag{4}$$

$$\begin{aligned}
\frac{\partial\mathcal{L}}{p_i^*} &= -y_i^T + p_i^* + \lambda\frac{p_i^*}{||p_i^*||_2} + \tau^* 1^T - \mu_i^* = 0 \\
&\Rightarrow (1+\frac{\lambda}{||p_i^*||_2})p_i^* = y_i - \tau^* + \mu_i^* \\
&\Rightarrow p_i^* = (1+\frac{\lambda}{||p_i^*||_2})^{-1}(y_i - \tau^* + \mu_i^*)_+
\end{aligned} \tag{5}$$

由于上式中存在$||p_i^*||_2$,因此将两面同时求l_2范数，解出$||p_i^*||_2$, 令$s_i = (y_i - \tau^* + \mu_i^*)_+$,其中$s_{ij} = (\theta_i^j - \tau^* + \mu_{ij}^*)_+$.

$$\begin{aligned}
||p_i^*||_2 &= ||(1+\frac{\lambda}{||p_i^*||_2})^{-1}s_i||_2 \\
||p_i^*||_2 &= \sqrt{(1+\frac{\lambda}{||p_i^*||_2})^{-2}\sum_{j=1}^{m} s_{ij}^2} \\
||p_i^*||_2 &= (1+\frac{\lambda}{||p_i^*||_2})^{-1}||s_i||_2 \\
||p_i^*||_2(1+\frac{\lambda}{||p_i^*||_2}) &= ||s_i||_2 \\
||p_i^*||_2 + \lambda &= ||s_i||_2 \\
||p_i^*||_2 &= ||s_i||_2 - \lambda
\end{aligned} \tag{6}$$

将求得的$||p_i^*||_2$带入Eq. 5:

$$\begin{aligned}
p_i^* &= (1+\frac{\lambda}{||s_i||_2})^{-1}s_i \\
p_i^* &= (1-\frac{\lambda}{||s_i||_2})s_i
\end{aligned} \tag{7}$$

2) 当$p_i = 0$时，由于导数不存在，因此尝试通过次梯度求解：令$f(p_i) = ||p_i||^2$,当$p_i = 0$时，

$$\begin{aligned}
\frac{\partial f(p_i)}{\partial p_i} &= \{v \in R^d | f(p_i^{'}) \le f(p_i) + v^T(p_i^{'} - p_i), \forall p_i^{'} \in R^d\} \\
&= \{v \in R^d | ||p_i^{'}||_2 \le v^T p_i^{'}, \forall p_i^{'} \in R^d\}
\end{aligned} \tag{8}$$

因此，p_i的次梯度向量v在$p_i = 0$时，需要满足$||v|| \leq 1$. 并且KKT条件要求$0 \in -y_i + \lambda v + \tau - \mu_i$，所以我们可以得到：

$$\begin{aligned} -y_i + \lambda v + \tau - \mu_i = 0 \\ \lambda v = y_i - \tau + \mu_i \\ v = \frac{1}{\lambda}(y_i - \tau + \mu_i) \\ v = \frac{1}{\lambda} s_i \end{aligned} \tag{9}$$

由于$||v||_2 \leq 1$, 所以得到:

$$\begin{aligned} ||\frac{1}{\lambda} s_i||_2 \leq 1 \\ ||s_i||_2 \leq \lambda \end{aligned} \tag{10}$$

所以，当$||s_i||_2 \leq \lambda$ 时，$p_i = 0$。

通过上面的分析可以得出p_i的解的形式为：

$$\begin{aligned} p_i &= (1 - \frac{\lambda}{||s_i||_2})_+ s_i \\ &= (1 - \frac{\lambda}{||(y_i - \tau + \mu_i)_+||_2})_+ (y_i - \tau + \mu_i)_+ \end{aligned} \tag{11}$$

进一步由Eq. 4可知，当$p_i > 0$时$\mu_i = 0$,因此，可化简为：

$$p_i = (1 - \frac{\lambda}{||(y_i - \tau)_+||_2})_+ (y_i - \tau)_+ \tag{12}$$