Link to slides (meeting notes): MR Fairness with Column Generation - Michael (Google Docs)

1 Problem Formulation

1.1 Master problem

$$\textbf{min} \quad \sum_{k} r_k z_k \tag{1a}$$

$$\textbf{s.t.} \quad \sum_{k \in C} s_{i,k} z_k = 1, \forall i \tag{1b}$$

$$\sum_{k \in C} t_{g,k} z_k \geq \beta_g, \forall g \tag{1c}$$

$$\sum_{k \in C} z_k = K \tag{1d}$$

$$z_k \in \{0, 1\} \tag{1e}$$

k: the k-th cluster of points
g: the group of points with label g
C: the set of all clusters
r_k : sum of squared 2-norm of distances between points in cluster k and the center of cluster k.
z_k : decision variables that determine if a cluster is selected.
β_g: least number of α represented cluster for group g.
$s_{i,k}$: parameters that determine if point i is in cluster k.
$t_{g,k}$: parameters that determine if cluster k is α-represented for group g.

1.2 Pricing problem

$$\textbf{min} \quad (\sum_{\forall i} r_i - \sum_{\forall i} \mu_i s_i - \sum_{\forall g} \lambda_g t_g) \tag{2a}$$

$$\textbf{s.t.} \quad \sum_{\forall i} q_{g,i} s_i + M(1 - t_g) \geq \alpha_g \sum_{\forall i} s_i, \forall g \tag{2b}$$

$$r_i + M(1 - s_i) \geq ||x_i - c||_2^2, \forall i \tag{2c}$$

$$u \geq \sum_{\forall i} s_i \geq l, \forall i \tag{2d}$$

$$r_i \geq 0, s_i, t_g \in \{0, 1\}, c \in \mathbb{R}^n \tag{2e}$$

s_i : decision variables determining if point i is in the generated cluster.
t_g : decision variables determining if the generated cluster is α-represented for group g.
r_i : decision variables for the contribution of squared 2-norm of point i to the sum of distances between points and the center of cluster in the new cluster.
c : decision variable for the center of the new cluster.
μ_i : dual variables for constraints 1b.
λ_g : dual variables for constraints 1c.
$q_{g,i}$: parameters that determine if point i is in group g.
u : upper bound for the number of points in a cluster
l : lower bound for the number of points in a cluster

1.3 Integer Program for solving an initial feasible set of clusters for the master problem

This problem is almost the same as the integer program proposed in Fair Minimum Representation Clustering, except that the objective can be arbitrary since we are only looking for a feasible solution.

$$
\begin{aligned}
\textbf{min} \quad & 0 && && (3)\\
\textbf{s.t.} \quad & \sum_{k \in \mathcal{K}} z_{ik} = 1 && \forall x^i \in \mathcal{X} && (4)\\
& \sum_{x^i \in X_g} z_{ik} + M(1 - y_{gk}) \geq \alpha \sum_{x^i \in \mathcal{X}} z_{ik} && \forall (g,k) \in \mathcal{W} && (5)\\
& \sum_{k \in \mathcal{K} : (g,k) \in \mathcal{W}} y_{gk} \geq \beta_g && \forall g \in \mathcal{G} && (6)\\
& l \leq \sum_{x^i \in \mathcal{X}} z_{ik} \leq u && \forall k \in \mathcal{K} && (7)\\
& z_{ik}, y_{gk} \in \{0, 1\} && \forall x^i \in \mathcal{X}, (g,k) \in \mathcal{W} && (8)
\end{aligned}
$$