1. Appendix

Here we present a more detailed proof of Theorem 1 and 2.

1.1. Proof of Theorem 1

We prove a more general result:

**Theorem 1.** Consider vectors \( x_i \in \mathbb{R}^m \) for \( i = 1, 2, \ldots, n \) and their partitions \( V_1, V_2, \ldots, V_K \) with sizes \( n_1, n_2, \ldots, n_K \). Take the SON optimization:

\[
\min_{\{u_i \in \mathbb{R}^m\}} \frac{1}{2} \sum_{i=1}^{n} \|x_i - u_i\|_2^2 + \lambda \sum_{i \neq j} \|u_i - u_j\|_2
\]

(1)

and its associated centroid optimization:

\[
\min_{\{v_{\alpha} \in \mathbb{R}^m, \alpha = 1, \ldots, K\}} \frac{1}{2} \sum_{i=1}^{K} \|v_{\alpha} - c_{\alpha}\|_2^2 + \lambda \sum_{\alpha \neq \beta} n_{\alpha}n_{\beta}\|c_{\alpha} - c_{\beta}\|_2
\]

(2)

where

\[
c_{\alpha} = \frac{\sum_{i \in V_{\alpha}} x_i}{n_{\alpha}}
\]

1. Suppose that for every \( \alpha \in [K] \),

\[
\max_{i, j \in V_{\alpha}} \|x_i - x_j\| \leq \lambda
\]

Then, \( u_i = v_{\alpha} \) for \( i \in V_{\alpha} \) is a global solution of the SON clustering.

2. If all \( c_{\alpha} \)s are distinct and \( d \geq \frac{1}{2n(K - 1)} \), then all centroids \( v_{\alpha} \) are distinct.

3. If \( \max \|c_{\alpha} - c_{\beta}\| \geq \lambda \) where \( c = \sum_{i=1}^{n} x_i/n \), then at least two centroids \( v_{\alpha} \) are distinct.

**Proof.** Notice that the solution of the centroid optimization satisfies

\[
c_{\alpha} - v_{\alpha} = \lambda \sum_{\beta} n_{\beta} z_{\alpha, \beta}
\]

where \( \|z_{\alpha, \beta}\| \leq 1 \), \( z_{\alpha, \beta} = -z_{\beta, \alpha} \) and whenever \( v_{\alpha} \neq v_{\beta} \), the relation \( z_{\alpha, \beta} = \frac{v_{\alpha} - v_{\beta}}{\|v_{\alpha} - v_{\beta}\|_2} \) holds. Now, for the solution \( u_i = v_{\alpha} \) for \( i \in V_{\alpha} \), define

\[
z_{i,j}' = \begin{cases} 
z_{\alpha, \beta} & \alpha \neq \beta \\
x_i - x_j & \alpha = \beta
\end{cases}
\]

where \( i \in V_{\alpha}, j \in V_{\beta} \). It is easy to see that \( \|z_{i,j}'\|_2 \leq 1 \), \( z_{i,j}' = -z_{j,i}' \) and whenever \( u_i \neq u_j \), we have that \( z_{i,j}' = \frac{u_i - u_j}{\|u_i - u_j\|_2} \). Further for each \( i \),

\[
\lambda \sum_{j} z_{i,j}' = \lambda \sum_{\beta} z_{\alpha, \beta} n_{\beta} + \sum_{j \in V_{\alpha}} \frac{x_i - x_j}{n_{\alpha}}
\]

\[
= c_{\alpha} - v_{\alpha} + x_i - c_{\alpha} = x_i - v_{\alpha} = x_i - u_i
\]

This shows that the local optimality conditions for the SON optimization holds and proves part a.

For part b, denote the solution of the centroid optimization by \( v_{\alpha}(\lambda) \) and notice that the solution of SON consists of distinct elements \( v_{\alpha} = c_{\alpha} \) and is continuous at \( \lambda = 0 \). Hence, \( v_{\alpha} \)s remain distinct in an interval \( \lambda \in (0, \lambda_1) \). Take \( \lambda_0 \) as the supremum of all possible \( \lambda_1 \)s. Hence, the solution in \( \lambda \in [0, \lambda_0] \) contains distinct element and at \( \lambda = \lambda_0 \) contains two equal elements (otherwise, one can extend \( \lambda \) to some \( [0, \lambda_0 + \epsilon] \), which is against \( \lambda \) being supremum). Now, notice that for \( \lambda \in [0, \lambda_0] \) the objective function is smooth at the optimal point. Hence, \( v_{\alpha}(\lambda) \) is differentiable and satisfies

\[
\delta = \left[ \frac{dv_{\alpha}}{d\lambda} \right] = H^{-1} \frac{\partial g}{\partial \lambda}
\]

(3)

where \( [\cdot]_\alpha \) and \( [\cdot]_{\alpha, \beta} \) denote block vectors and block matrices respectively. Moreover, \( H \) and \( g \) are the Hessian and the gradient of the objective function at the optimal point. In other words,

\[
H = \left[ \begin{array}{c}
\alpha \\
\beta
\end{array} \right]_{\alpha, \beta} n_{\alpha} n_{\beta}
\]

and

\[
\frac{\partial g}{\partial \lambda} = \left[ \begin{array}{c}
\alpha \\
\beta
\end{array} \right] \sum_{\alpha} z_{\alpha, \beta} n_{\alpha} n_{\beta}
\]

Hence,

\[
\delta = \left[ \delta_{\alpha, \beta} + I \|v_{\alpha} - v_{\beta}\|_2^2 - (v_{\alpha} - v_{\beta}) (v_{\alpha} - v_{\beta})^T \lambda n_{\beta} \right]_{\alpha, \beta}^{-1}
\]

\[
\times \left[ \begin{array}{c}
\alpha \\
\beta
\end{array} \right] \sum_{\beta} z_{\alpha, \beta} n_{\beta}
\]

Simple calculations show that \( \|\delta\|_2 \leq n \sqrt{K} \). Hence,

\[
\left\| \frac{dv_{\alpha}}{d\lambda} \right\|_2 \leq \|\delta\|_2 \leq \sqrt{Kn}
\]

This yields for \( \lambda < \lambda_0 \) to

\[
\|v_{\alpha}(\lambda) - v_{\beta}(\lambda)\|_2 = \left\| c_{\alpha} - c_{\beta} + \int \left( \frac{dv_{\alpha}}{d\lambda} - \frac{dv_{\beta}}{d\lambda} \right) d\lambda \right\|_2
\]

\[
\geq \|c_{\alpha} - c_{\beta}\|_2 - \int \left\| \frac{dv_{\alpha}}{d\lambda} - \frac{dv_{\beta}}{d\lambda} \right\|_2 d\lambda
\]
Theorem 2. Starting from $v_0 = v_\beta$ for some $\alpha \neq \beta$, we get that $d - 2n\lambda \sqrt{K} \leq 0$ or $\lambda_0 \geq d/2n\sqrt{K}$, this proves part b.

For part c, take a value of $\lambda$, where $v_1 = v_2 = \ldots = v_K$. It is simple to see that in this case $v_\alpha = c$. The optimality condition leads to

$$c - c_\alpha = \lambda \sum_{\beta \neq \alpha} z_{\alpha, \beta} n_\beta$$

Hence, $\|c - c_\alpha\|_2 \leq \lambda(n - n_\alpha)$. This proves part c.

\[ \Box \]

1.2. Proof of Theorem 2

Denote by $U_k$ a matrix where the $i^{th}$ column is the value of $u_i$ at the $k^{th}$ iteration. Define

$$\psi_\mu(U) = \mathcal{E}(U_{k+1} \mid U_k = U, \mu_k = \mu),$$

which by simple manipulations leads to

$$\psi_\mu(U) = U + \frac{1}{n} \sum_{i<j} \left( L_{ij}(\Pi_{ij}^\mu(u_i, u_j)) - L_{ij}(u_i, u_j) \right)$$

where $U$ denotes the $i^{th}$ column of $U$ and $L_{ij}(x,y)$ is a matrix where the $i^{th}$ column is $x$, the $j^{th}$ column is $y$ and the rest are zero. Also, denote

$$\sigma^2_\mu(U) = \operatorname{Var}(U_{k+1} \mid U_k = U, \mu_k = \mu) = \mathcal{E} \left( \|U_{k+1}\|_2^2 \mid U_k = U, \mu_k = \mu \right) - \|\phi_\mu(U)\|_2^2$$

We prove a more detailed theorem:

**Theorem 2.** Starting from $U_0 = U_0$ (the initialization of the algorithm), define the characteristic sequence $\{U_k\}_{k=0}^\infty$ by the following iteration:

$$U_{k+1} = \psi_\mu_k(U_k)$$

1. We have that

$$\Pr \left( \sup_k \|U_k - \bar{U}_k\|_F^2 + \sum_{l=k}^\infty \mu_l^2 > \lambda \right) \leq \frac{\sum_{k=0}^\infty \mu_k^2}{\lambda}$$

2. Define $\bar{U}$ as the unique optimal solution of the SON optimization and suppose that $\{\mu_k\}$ is a non-increasing sequence.

\(a\) There exists a positive sequence $h_n = O(1/n)$, where $n$ is the number of data points, such that

$$R(U_k, \mu_k) \leq h_n \sum_{l=0}^{k-1} \mu_l^2 e^{-\frac{2}{\lambda} \sum_{s=0}^{k-1} \mu_s} + R(U_0, \mu_0) e^{-\frac{2}{\lambda} \sum_{s=0}^{k-1} \mu_s}$$

where

$$R(U, \mu) = \frac{1}{2} \|U - \bar{U}\|^2_\Phi + \mu \left( \Phi(U) - \Phi(U) \right)$$

\(b\) There exists a universal constant $a$ such that

$$\|U_k - \bar{U}\|^2_\Phi \leq a \sum_{l=0}^{k-1} \mu_l^2 e^{-\frac{2}{\lambda} \sum_{s=0}^{k-1} \mu_s} + \|U_0 - \bar{U}\|^2_\Phi e^{-\frac{2}{\lambda} \sum_{s=0}^{k-1} \mu_s}$$

3. Assume that $\{\mu_k\}$ is non-increasing and $\sum_{l=0}^{\infty} \mu_l^2 < \infty$. Then, the sequence $U_k$ converges to $\bar{U}$ in the following strong probability sense:

$$\forall \epsilon > 0; \lim_{k \to \infty} \mathbb{P} \left( \sup_{l \geq k} \|U_l - \bar{U}\|^2_\Phi > \epsilon \right) = 0$$

4. Take $\mu_k = \frac{\mu_k}{k^\alpha}$ for $k = 1, 2, \ldots$ and $\frac{2}{3} < \alpha < 1$. For sufficiently small values of $\epsilon > 0$ the relation

$$\|U_l - \bar{U}\|^2_\Phi = O\left( \frac{1}{(1-\alpha)^2} \right)$$

holds with probability $1$.

**Proof.** Denote by $\Omega_k$ the pair $(i,j)$ which is selected in iteration $k$ and $\bar{\Omega} = (\Omega_0, \Omega_1, \ldots, \Omega_{k-1})$. Also, denote $\psi_\mu(U, (i,j)) = U + L_{ij}(\Pi_{ij}^\mu(u_i, u_j)) - L_{ij}(u_i, u_j)$. Then, the iterations can be written as

$$U_{k+1} = \psi_\mu_k(U_k, \Omega_k)$$

Define $\Delta_k = U_k - \bar{U}_k$ and $\eta_k = \psi_\mu_k(U_k, \Omega_k) - \mathcal{E}(\psi_\mu(U_k, \Omega_k) \mid \bar{U}_k)$. Also, denote $U = \{U_k\}_{k=0}^\infty$. Notice that the sequence $\{\eta_k\}_{k=0}^\infty$ consists of zero-mean independent elements. Subtracting the two iterations in (9) gives us:

$$\Delta_{k+1} = \psi_\mu_k(U_k, \Omega_k) - \psi_\mu_k(U_k, \Omega_k) + \eta_k$$

It is simple to see that $\Pi_{ij}^\mu(u_i, u_j)$ is a contraction map for any $\mu, i, j$. Then, it is simple to deduce that $\psi_\mu(U, \Omega)$ is
a contraction map for any $\Omega$ and $\mu$. As a result, we obtain from (10) that

$$E \left( \|\Delta_{k+1} - \eta_k\|_F^2 \mid \Omega^k \right) \leq \|\Delta_k\|_F^2,$$

which can also be written as

$$E \left( \|\Delta_{k+1}\|_F^2 \mid \Omega^k \right) \leq \|\Delta_k\|_F^2 + 2E \left( \langle \psi_{\mu}(U_k, \eta_k) \rangle \mid \Omega^k \right) - E\|\eta_k\|_F^2,$$

Now, it is simple to see that $\|\psi_{\mu}(U, \Omega) - U\| \leq \sqrt{2\mu}$. Furthermore, $U_k$ only depends on $\Omega_0, \Omega_1, \ldots, \Omega_{k-1}$, while $\eta_k$ is a function of $\Omega_k$. Hence, $U_k$ and $\eta_k$ are independent and $E\langle U_k, \eta_k \rangle = 0$. This leads to

$$E \left( \|\Delta_{k+1}\|_F^2 \mid \Omega^k \right) \leq \|\Delta_k\|_F^2 + 2E \left( \langle \psi_{\mu}(U_k, \Omega_k) - U_k, \eta_k \rangle \mid \Omega^k \right) - E\|\eta_k\|_F^2,$$

Notice that $E\|\eta_k\|_F^2 = \sigma_{\mu}^2(U_1)$ and

$$\|U_{k+1} - U_k\|_2 = \|\psi_{\mu}(U_k, \Omega_k) - U_k\|_2 \leq \sqrt{2\mu}.$$

which leads to

$$\sigma_{\mu}^2(U) \leq 2\mu^2.$$

We conclude that

$$E \left( \|\Delta_{k+1}\|_F^2 \mid \Omega^k \right) \leq \|\Delta_k\|_F^2 + 4\mu_k^2.$$

Define $s_k = \sum_{l=k}^\infty \mu_l^2$. We observe that $\|\Delta_k\|_F^2 + s_k$ is a supermartingale. Hence, from the supermartingale version of the Doob’s inequality we obtain that

$$\Pr \left( \sup_{k} \|\Delta_k\|_F^2 + s_k > \lambda \right) \leq \frac{E\|\Delta_0\|_F^2 + s_0}{\lambda} = \sum_{k=0}^\infty \frac{\mu_k^2}{\lambda}$$

This proves part (1).

For part (2) from the definition of the proximal operator, there exists a vector $\zeta \in \partial \phi_{\Omega}(\psi_{\mu}(U, \Omega))$ such that

$$\psi_{\mu}(U, \Omega) = U - \mu \zeta.$$

We conclude that

$$\phi_{\Omega}(U) - \phi_{\Omega}(\psi_{\mu}(U, \Omega)) \geq \frac{1}{\mu} \langle U - \psi_{\mu}(U, \Omega), \zeta - \psi_{\mu}(U, \Omega) \rangle = \frac{1}{2\mu} \left( \|U - \psi_{\mu}(U, \Omega)\|_F^2 - \|U - \psi_{\mu}(U, \Omega)\|_F^2 + \|U - \psi_{\mu}(U, \Omega)\|_F^2 \right).$$

Hence,

$$\Phi(U) - \phi_{\Omega}(\psi_{\mu}(U, \Omega)) \geq \frac{n(n-1)}{2\mu} \left( \|U - \psi_{\mu}(U, \Omega)\|_F^2 - \|U - \psi_{\mu}(U, \Omega)\|_F^2 \right) \geq \frac{n(n-1)}{2\mu} \left( \|U - \psi_{\mu}(U, \Omega)\|_F^2 - \|U - \psi_{\mu}(U, \Omega)\|_F^2 \right) \leq \|U - \psi_{\mu}(U, \Omega)\|_F^2$$

where the last inequality is obtained by Jensen’s inequality. Notice that

$$\sum_{\Omega} \phi_{\Omega}(\psi_{\mu}(U, \Omega)) =$$

$$\sum_{\Omega, \Omega'} \phi_{\Omega'}(\psi_{\mu}(U, \Omega)) - \sum_{\Omega \neq \Omega'} \phi_{\Omega'}(\psi_{\mu}(U, \Omega)) \geq \frac{n(n-1)}{2} \Phi(U) - \sum_{\Omega, \Omega'} \phi_{\Omega'}(\psi_{\mu}(U, \Omega))$$

$$= \Phi(U) + \frac{n(n-1)}{2} \left( \Phi(U) - \Phi(U) \right) - 2(n-2) \alpha \mu$$

Now, notice that $\phi_{\Omega'}(\psi_{\mu}(U, \Omega)) = \phi_{\Omega'}(U, \Omega) = 0$ when $\Omega$ and $\Omega'$ do not overlap. Also, there exists a constant $\alpha$ such that $|\phi_{\Omega'}(\psi_{\mu}(U, \Omega)) - \phi_{\Omega'}(\psi_{\mu}(U, \Omega))] < \alpha \mu$. We conclude that

$$\sum_{\Omega} \phi_{\Omega}(\psi_{\mu}(U, \Omega)) \geq$$

$$\Phi(U) - \frac{n(n-1)}{2} \left( \Phi(U) - \Phi(U) \right) - 2(n-2) \alpha \mu$$

Define $\eta_1 = 8(n-2)\alpha/n(n-1) = O(\frac{1}{n})$. Replacing this result in (11) and performing straightforward calculations leads to

$$h_{n, \mu}^2 \geq \frac{2\sqrt{\mu}}{n^2} \left( \Phi(U) - \Phi(\tilde{U}) \right)$$

$$+ \mu (\Phi(U) - \Phi(U))$$

$$+ \frac{1}{2} \left( \|U - \psi_{\mu}(U)\|_F^2 - \|\tilde{U} - \psi_{\mu}(U)\|_F^2 \right)$$

Now, we introduce the recursion to (12). We introduce $R_k = R(U_k, \mu_k)$ and use monotonicity of $\mu_k$ to conclude that:

$$h_{n, \mu_k}^2 \geq \frac{2\mu_k}{n^2} \left( \Phi(U_k) - \Phi(\tilde{U}) \right) + R_{k+1} - R_k$$

Finally, we use the fact that $\Phi(\cdot)$ is a $1-$strongly convex function which leads to $\Phi(U) - \Phi(\tilde{U}) \geq \frac{1}{2} \|U - \tilde{U}\|_F^2$, and conclude that

$$\Phi(U) - \Phi(\tilde{U}) \geq \frac{R(U, \mu)}{1 + \mu}$$

This yields to

$$R_{k+1} - h_{n, \mu_k}^2 \leq \left( 1 - \frac{2\mu_k}{n(n-1)} \right) R_k \leq e^{-\frac{2\mu_k}{n(n-1)}} R_k$$

where the last equality holds because $1 - x \leq e^{-x}$ for every positive $x$. It is now simple to see by induction that

$$R_k \leq h_n \sum_{l=0}^{k-1} \mu_l^2 e^{-\frac{2\mu_l}{n(n-1)}} + R_0 e^{-\frac{2\mu_0}{n(n-1)}} \sum_{l=0}^{k-1} \mu_l^2$$

(13)
which proves part (2a).

For part (2b), we observe from (11) that

\[
\Phi(\tilde{U}) - \Phi(U) + \frac{n(n-1)}{2} a\mu \geq \frac{n(n-1)}{4\mu} \left( \|\tilde{U} - \psi_\mu(U)\|_F^2 - \|\tilde{U} - U\|_F^2 \right)
\]

which with the similar argument to above leads to

\[
\frac{1}{2} \|\tilde{U}_{k+1} - \tilde{U}\|_F^2 \leq \left( 1 - \frac{2\mu_k}{m^2} \right) \frac{1}{2} \|\tilde{U}_k - \tilde{U}\|_F^2 + a\mu_k^2
\]

\[
\leq \frac{1}{2} \|\tilde{U}_k - \tilde{U}\|_F^2 \frac{2\mu_k}{m^2} + a\mu_k^2
\]

We conclude part (2b).

For part (3), define \( l^k \) as \( \{U^k_i\}_{i=0}^\infty \), as the sequence obtained by starting from \( U^k_0 = U_k \) and applying

\[
\tilde{U}^k_{l+1} = \psi_{\mu_{l+1}}(\tilde{U}^k_l)
\]

Take arbitrary (non-zero) positive numbers \( \epsilon, \delta \). Take \( \lambda \) such that \( \lambda \geq \frac{3}{2} \sum_{i=l}^{\infty} \mu_i^2 \). Define

\[
\Phi_{\text{max}} = \max_{\|U - U\| \leq \lambda} \Phi(U)
\]

Define \( l_0, k \) such that \( \sum_{i=k}^{\infty} \mu_i^2 < \epsilon \delta/8 \) and

\[
\forall l > l_0: h_n \sum_{i=0}^{l-1} \mu_{t+k} e^{-\frac{\mu_{t+k}}{m^2}} + \left( \lambda + \mu_k \Phi_{\text{max}} \right) e^{-\frac{\mu_{t+k}}{m^2}} < \frac{\epsilon}{8}
\]

It is simple to see that such a choice exists because of the conditions in part (3). Now, we define two outcomes \( H_1 \) and \( H_2 \):

\[
H_1 : \forall k \geq 0: \|U_k - \tilde{U}\|_F^2 \leq \lambda \\
H_2 : \forall l \geq 0: \|\tilde{U}_k^l - U_{l+k}\| \leq \epsilon
\]

Notice that from part (1) we have that \( \text{Pr}(H_1^c) \) and \( \text{Pr}(H_2^c) \) are less than \( \delta/2 \). Furthermore, under \( H_1 \cap H_2 \) we have that:

\[
\forall l > l_0: \|U_{l+k} - \tilde{U}\|_F^2 \leq 2(\|U_{l+k} - \tilde{U}_l^f\|_F^2 + \|\tilde{U}_l^f - \tilde{U}\|_F^2)
\]

\[
\leq 2\left( \frac{\epsilon}{4} + \frac{\epsilon}{4} \right) = \epsilon
\]

This is because according to part (2),

\[
\|\tilde{U}^f_k - U\|_F^2 \leq 2R(\tilde{U}^f_k, \mu_{l+k}) \leq
\]

\[
2h_n \sum_{t=0}^{l-1} \mu_{t+k} e^{-\frac{\mu_{t+k}}{m^2}} + \frac{2}{\sum_{i=l+1}^{l-1} \frac{\mu_{t+k}}{m^2}} \leq \epsilon
\]

and

\[
+2R(U_k, \mu_k) e^{-\frac{\mu_k}{m^2}} \leq \frac{\epsilon}{4}
\]

where we used \( H_1 \) to conclude that \( R(U_k, \mu_k) \leq \lambda + \Phi_{\text{max}}\mu_k \). We conclude that

\[
\text{Pr}( l_{t+1} > l_0 + k, \|U_t - \tilde{U}\|_F^2 > \epsilon ) \leq \text{Pr}(H_1^c) + \text{Pr}(H_2^c) \leq \delta
\]

which proves part (3).

For part (4), define \( k_r = r^\gamma, \lambda_r = r^{-\beta} \), where \( \gamma = \frac{1-\delta}{\alpha}, \beta < 2(\alpha - 1) - 1 \), and the outcomes:

\[
Q_r : \sup_{l \geq 0} ||U_{l+k_r}-\tilde{U}^k_l||_F^2 > \lambda_r
\]

By part (1), we have that

\[
\sum_{r=1}^{\infty} \text{Pr}(Q_r) < \infty.
\]

Hence by Borel-Cantelli lemma, \( Q_r, Q_{r+1}, Q_{r+2}, \ldots \) simultaneously hold for some \( r_0 \) with probability 1. For simplicity and without loss of generality, we assume that \( r_0 = 0 \) as it does not affect the asymptotic rate. Then for any \( r > 0 \), we have that

\[
\sup_{l \geq 0} ||U_{l+k_r}-\tilde{U}^k_l||_F^2 \leq \lambda_r
\]

In particular,

\[
||U_{l+k_r}-\tilde{U}^k_l||_F^2 \leq \lambda_r
\]

where \( k_r = k_{r+1} - k_r \). From part (2b), we conclude that

\[
||U_{l_r+k_r}-\tilde{U}^k_{l_r}||_F^2 \leq A \sum_{t=0}^{l_r-1} \frac{1}{(t+k_r)^{2\alpha}} e^{-\frac{\mu}{m^2}} + \frac{1}{(t+k_r)^{2\alpha}} \leq \lambda
\]

where we introduce \( \mu_1 = bn^2 \) and \( A = 4an^4b^2 \) for simplicity. This leads to

\[
||U_{l_r+k_r}-\tilde{U}^k_{l_r}||_F^2 \leq 2\lambda_r + A \sum_{t=0}^{l_r-1} \frac{1}{(t+k_r)^{2\alpha}} e^{-\frac{\mu}{m^2}} + \frac{1}{(t+k_r)^{2\alpha}} \leq \lambda
\]

\[
+2\|U_{k_r} - \tilde{U}^k_{l_r}||_F^2 \leq \sum_{t=0}^{l_r-1} \frac{1}{(t+k_r)^{2\alpha}} e^{-\frac{\mu}{m^2}} + \sum_{t=0}^{l_r-1} \frac{1}{(t+k_r)^{2\alpha}} e^{-\frac{\mu}{m^2}} \leq \lambda
\]

\[
\leq L l_r^{(k_r-\alpha) - \alpha - l_r} \|U_{k_r} - \tilde{U}^k||_F^2
\]

\[
2\lambda_r + A \sum_{t=0}^{l_r} \frac{1}{(t+k_r)^{2\alpha}} e^{(l_r+k_r)1-\alpha-l_r} \leq \lambda
\]
where $L$ denotes "some suitable constant" which may vary in difference occurrences. Notice that

$$
\sum_{t=0}^{l_r} \frac{1}{(t + k_r)^{2\alpha}} e^{L(k_r + t)^{1-\alpha} - Lk_r^{1-\alpha}} = \sum_{t=k_r}^{k_{r+1}} \frac{1}{t^{2\alpha}} e^{L_t^{1-\alpha} - Lk_r^{1-\alpha}}
$$

$$
\leq L \sum_{t=k_r}^{k_{r+1}} \frac{1}{t^{2\alpha}} e^{-L\rho \log(k_{r+1})}
$$

$$
+ \sum_{t=k_{r+1}}^{k_{r+1} - Lk_r^{\alpha}(1 + \rho \log(k_{r+1}))} \frac{1}{t^{2\alpha}} e^{-L\rho \log(k_{r+1})}
$$

$$
\leq L \left( \frac{1}{(k_{r+1} - Lk_r^{\alpha}(1 + \rho \log(k_{r+1}))^{2\alpha} - 1)} \right)
$$

$$
+ Le^{-L\rho \log(k_{r+1})} \leq \frac{L \log(k_{r+1})}{k_r^{\alpha}} \leq \frac{L \log r}{r^{\gamma \alpha}} < \frac{L}{r^\beta}
$$

where $\rho$ is a sufficiently large constant and we use the fact that $\gamma \alpha > \gamma(2\alpha - 1) - 1 > \beta$. Moreover,

$$
k_r^{1-\alpha} - k_r^{1-\alpha} = r^{\gamma(1-\alpha)} - (r+1)^{\gamma(1-\alpha)} \leq -Lr^{\gamma(1-\alpha)} - 1
$$

We conclude that

$$
\|U_{k_{r+1}} - \tilde{U}\|_F^2 \leq \frac{L}{r^\beta} + Le^{-Lr^{\gamma(1-\alpha)} - 1} \|U_{k_r} - \tilde{U}\|_F^2
$$

which leads to

$$
\|U_{k_r} - \tilde{U}\|_F^2 \leq L \left( \sum_{i=1}^{r-1} \frac{1}{s^\beta} e^{-L \sum_{i=1}^{r-1} e^{\gamma(1-\alpha)}} + e^{-L \sum_{i=1}^{r-1} e^{\gamma(1-\alpha)}} \right)
$$

$$
\leq L \left( \sum_{i=1}^{r-1} \frac{1}{s^\beta} e^{L(\gamma(1-\alpha) - r^{\gamma(1-\alpha)})} + e^{-Lr^{\gamma(1-\alpha)}} \right)
$$

With a similar approach to the above, we observe that

$$
\|U_{k_r} - \tilde{U}\|_F^2 \leq \frac{L \log r}{r^{\beta - \epsilon}} \leq \frac{L}{r^{\beta - \epsilon}}
$$

Take $k_r < l \leq k_{r+1}$. We observe that

$$
\|U_l - \tilde{U}\|_2^2 \leq 2(\|U_{k_r} - \tilde{U}\|_2^2 + \|U_{k_r} - U_l\|_2^2)
$$

$$
\leq 2\lambda_r + \frac{L}{r^{\beta - \epsilon}} \leq \frac{L}{l^{\gamma - \epsilon}}
$$

By taking $\beta = \gamma(2\alpha - 1) - 1$, we obtain part (4).