Appendix A
Proofs

A.1 Proof of Theorem 4.1:

Proof. For 1), let X^* denote the optimal k-means centers of P. Since $\mathbf{S}$ is an ϵ-coreset with probability $\geq 1-\delta$ (Theorem 3.2), by Definition 3.2, the following holds with probability $\geq 1-\delta$:

$$\begin{aligned}\text{cost}(P,X) &\leq \frac{1}{1-\epsilon}\text{cost}(\mathbf{S},X) \leq \frac{1}{1-\epsilon}\text{cost}(\mathbf{S},X^*)\\ &\leq \frac{1+\epsilon}{1-\epsilon}\text{cost}(P,X^*). \quad (25)\end{aligned}$$

For 2), the cost of transferring (S,Δ,w) is dominated by the cost of transferring S. Since S lies in a d'-dimensional subspace spanned by the columns of $V^{(d')}$, it suffices to transmit the coordinates of each point in S in this subspace together with $V^{(d')}$. The former incurs a cost of $O(|S|\cdot d')$, and the latter incurs a cost of $O(dd')$. Plugging $d'=O(k/\epsilon^2)$ and $|S|=\tilde{O}(k^3/\epsilon^4)$ from Theorem 3.2 yields the overall communication cost as $O(dk/\epsilon^2)$. ☐

A.2 Proof of Lemma 4.1

Proof. Let $\delta'=\delta/(2nk)$. By the JL Lemma (Lemma 3.1), there exists $d'=O(\epsilon^{-2}\log(1/\delta'))=O(\epsilon^{-2}\log(nk/\delta))$, such that every $x\in\mathbb{R}^d$ satisfies $\|\pi(x)\|\approx_{1+\epsilon}\|x\|$ with probability $\geq 1-\delta'$. By the union bound, this implies that with probability $\geq 1-\delta$, every $p-x_i$ for $p\in P$ and $x_i\in X\cup X^*$ satisfies $\|\pi(p)-\pi(x_i)\|\approx_{1+\epsilon}\|p-x_i\|$. Therefore, with probability $\geq 1-\delta$,

$$\begin{aligned}\text{cost}(\pi(P),\pi(X)) &= \sum_{p\in P}\min_{x_i\in X}\|\pi(p)-\pi(x_i)\|^2 \quad (26)\\ &\leq \sum_{p\in P}\min_{x_i\in X}(1+\epsilon)^2\|p-x_i\|^2\\ &= (1+\epsilon)^2\text{cost}(P,X), \quad (27)\end{aligned}$$

$$\begin{aligned}(26) &\geq \sum_{p\in P}\min_{x_i\in X}\frac{1}{(1+\epsilon)^2}\|p-x_i\|^2\\ &= \frac{1}{(1+\epsilon)^2}\text{cost}(P,X). \quad (28)\end{aligned}$$

Combining (27) and (28) proves (5). Similar argument will prove (6). ☐

A.3 Proof of Theorem 4.2

Proof. For 1), let X^* be the optimal k-means centers of P and $\mathbf{S}':=(S',\Delta,w)$ be generated in line 3 of Algorithm 1. With probability at least $(1-\delta)^2$, π_1 satisfies (5, 6) and π_2 generates an ϵ-coreset. Thus, with probability at least $(1-\delta)^2$,

$$\begin{aligned}\text{cost}(P,X) &\leq (1+\epsilon)^2\text{cost}(\pi_1(P),X') \quad (29)\\ &\leq \frac{(1+\epsilon)^2}{1-\epsilon}\text{cost}(\mathbf{S}',X') \quad (30)\\ &\leq \frac{(1+\epsilon)^2}{1-\epsilon}\text{cost}(\mathbf{S}',\pi_1(X^*)) \quad (31)\\ &\leq \frac{(1+\epsilon)^3}{1-\epsilon}\text{cost}(\pi_1(P),\pi_1(X^*)) \quad (32)\\ &\leq \frac{(1+\epsilon)^5}{1-\epsilon}\text{cost}(P,X^*), \quad (33)\end{aligned}$$

where (29) is by (5) and that $\pi_1(X)=X'$, (30) is because $\mathbf{S}'$ is an ϵ-coreset of $\pi_1(P)$, (31) is because X' minimizes $\text{cost}(\mathbf{S}',\cdot)$, (32) is again because $\mathbf{S}'$ is an ϵ-coreset of $\pi_1(P)$, and (33) is by (6).

For 2), the communication cost is dominated by transmitting S'. By Lemma 4.1, the dimension of P' is $d'=O(\epsilon^{-2}\log(nk/\delta))=O(\epsilon^{-2}\log n)$. By Theorem 3.2, the cardinality of S' is $|S'|=O(k^3\epsilon^{-4}\log^2(k)\log(1/\delta))$. Moreover, points in S' lie in a $\tilde{d}$-dimensional subspace for $\tilde{d}=O(k/\epsilon^2)$. Thus, it suffices to transmit the coordinates of points in S' in the $\tilde{d}$-dimensional subspace and a basis of the subspace. Thus, the total communication cost is

$$\begin{aligned}O((|S'|+d')\tilde{d}) &= O\left(\frac{k^4}{\epsilon^6}\log^2(k)\log(\frac{1}{\delta})+\frac{k}{\epsilon^4}\log n\right)\\ &= O\left(\frac{k\log n}{\epsilon^4}\right). \quad (34)\end{aligned}$$

For 3), note that for a given projection matrix $\Pi\in\mathbb{R}^{d\times d'}$ such that $\pi_1(P):=A_P\Pi$, line 2 takes $O(ndd')=O(nd\epsilon^{-2}\log n)$ time, where we have plugged in $d'=O(\epsilon^{-2}\log n)$. By Theorem 3.2, line 3 takes time

$$\begin{aligned}&O\left(\min(nd'^2,n^2d')+\frac{nk}{\epsilon^2}(d'+k\log(\frac{1}{\delta}))\right)\\ =&O\left(\frac{n}{\epsilon^2}(\frac{\log^2 n}{\epsilon^2}+\frac{k\log n}{\epsilon^2}+k^2\log(\frac{1}{\delta}))\right). \quad (35)\end{aligned}$$

Thus, the total complexity at the data source is:

$$\begin{aligned}&O\left(\frac{n}{\epsilon^2}(\frac{\log^2 n}{\epsilon^2}+\frac{k\log n}{\epsilon^2}+d\log n+k^2\log(\frac{1}{\delta}))\right)\\ =&O\left(\frac{nd}{\epsilon^2}\log^2 n\right)=\tilde{O}\left(\frac{nd}{\epsilon^2}\right). \quad (36)\end{aligned}$$

☐

A.4 Proof of Lemma 4.2

Proof. The proof is analogous to that of Lemma 4.1. Let $\delta'=\delta/(2n'k)$. Then there exists $d'=O(\epsilon^{-2}\log(1/\delta'))=O(\epsilon^{-2}\log(n'k/\delta))$, such that every $x\in\mathbb{R}^d$ satisfies $\|\pi(x)\|\approx_{1+\epsilon}\|x\|$ with probability $\geq 1-\delta'$. By the union bound, this implies that with probability $\geq 1-\delta$, every $p\in S$ and $x_i\in X\cup X^*$ satisfy $\|\pi(p)-\pi(x_i)\|\approx_{1+\epsilon}\|p-x_i\|$. Therefore,

$$\begin{aligned}&\text{cost}((\pi(S),\Delta,w),\pi(X))\\ &=\sum_{p\in S}w(p)\cdot\min_{x_i\in X}\|\pi(p)-\pi(x_i)\|^2+\Delta \quad (37)\\ &\leq(1+\epsilon)^2\left(\sum_{p\in S}w(p)\cdot\min_{x_i\in X}\|p-x_i\|^2+\Delta\right)\\ &=(1+\epsilon)^2\text{cost}(\mathbf{S},X), \quad (38)\end{aligned}$$

$$\begin{aligned}(37) &\geq \frac{1}{(1+\epsilon)^2}\left(\sum_{p\in S}w(p)\cdot\min_{x_i\in X}\|p-x_i\|^2+\Delta\right)\\ &=\frac{1}{(1+\epsilon)^2}\text{cost}(\mathbf{S},X), \quad (39)\end{aligned}$$

which prove (7). Similar argument proves (8). ☐

A.5 Proof of Theorem 4.3

Proof. For 1), let X^* be the optimal k-means centers of P and $\mathbf{S} := (S, \Delta, w)$ generated in line 2 of Algorithm 2. With probability at least $(1-\delta)^2$, $\mathbf{S}$ is an ϵ-coreset of P, and π_1 satisfies (7, 8). Thus, with this probability,

$$\begin{aligned}
\text{cost}(P, X) &\le \frac{1}{1-\epsilon}\text{cost}(\mathbf{S}, X) && (40)\\
&\le \frac{(1+\epsilon)^2}{1-\epsilon}\text{cost}((S', \Delta, w), X') && (41)\\
&\le \frac{(1+\epsilon)^2}{1-\epsilon}\text{cost}((S', \Delta, w), \pi_1(X^*)) && (42)\\
&\le \frac{(1+\epsilon)^4}{1-\epsilon}\text{cost}(\mathbf{S}, X^*) && (43)\\
&\le \frac{(1+\epsilon)^5}{1-\epsilon}\text{cost}(P, X^*), && (44)
\end{aligned}$$

where (40) is because $\mathbf{S}$ is an ϵ-coreset of P, (41) is due to (7) (note that $\pi_1(S) = S'$ and $\pi_1(X) = X'$), (42) is because X' minimizes $\text{cost}((S', \Delta, w), \cdot)$, (43) is due to (8), and (44) is again because $\mathbf{S}$ is an ϵ-coreset of P.

For 2), note that by Theorem 3.2, the cardinality of S needs to be $n' = O(k^3\epsilon^{-4}\log^2(k)\log(1/\delta))$. By Lemma 4.2, the dimension of S' needs to be $d' = O(\epsilon^{-2}\log(n'k/\delta))$. Thus, the cost of transmitting (S', Δ, w), dominated by the cost of transmitting S', is

$$\begin{aligned}
O(n'd') &= O\left(\frac{k^3\log^2 k}{\epsilon^6}\log(\frac{1}{\delta})\Big(\log k + \log(\frac{1}{\epsilon}) + \log(\frac{1}{\delta})\Big)\right)\\
&= \tilde{O}\left(\frac{k^3}{\epsilon^6}\right). && (45)
\end{aligned}$$

For 3), we know from Theorem 3.2 that line 2 of Algorithm 2 takes time $O(\min(nd^2, n^2d) + nk\epsilon^{-2}(d + k\log(1/\delta)))$. Given a projection matrix $\Pi \in \mathbb{R}^{d\times d'}$ such that $\pi_1(S) := A_S\Pi$, line 3 takes time $O(n'dd')$. Thus, the total complexity at the data source is

$$\begin{aligned}
&O\left(\min(nd^2, n^2d) + \frac{k}{\epsilon^2}nd + \frac{k^2\log k}{\epsilon^2}n + \frac{k^3\log^3 k(\log k + \log(\frac{1}{\epsilon}))}{\epsilon^6}d\right)\\
&= O\left(nd\cdot\min(n, d)\right). && (46)
\end{aligned}$$

□

A.6 Proof of Theorem 4.4

Proof. Let $n' := |S|$, d' be the dimension after $\pi_1^{(1)}$, and d'' be the dimension after $\pi_1^{(2)}$. Let X^* be the optimal k-means centers for P.

For 1), note that with probability $\ge (1-\delta)^3$, $\pi_1^{(1)}$ and $\pi_1^{(2)}$ will preserve the k-means cost up to a multiplicative factor of $(1+\epsilon)^2$, and π_2 will generate an ϵ-coreset of P'. Thus, with this probability, we have

$$\begin{aligned}
\text{cost}(P, X) &\le (1+\epsilon)^2\text{cost}(P', \pi_1^{(1)}(X)) && (47)\\
&\le \frac{(1+\epsilon)^2}{1-\epsilon}\text{cost}((S, \Delta, w), \pi_1^{(1)}(X)) && (48)\\
&\le \frac{(1+\epsilon)^4}{1-\epsilon}\text{cost}((S', \Delta, w), \pi_1^{(2)}\circ\pi_1^{(1)}(X)) && (49)\\
&\le \frac{(1+\epsilon)^4}{1-\epsilon}\text{cost}((S', \Delta, w), \pi_1^{(2)}\circ\pi_1^{(1)}(X^*)) && (50)\\
&\le \frac{(1+\epsilon)^6}{1-\epsilon}\text{cost}((S, \Delta, w), \pi_1^{(1)}(X^*)) && (51)\\
&\le \frac{(1+\epsilon)^7}{1-\epsilon}\text{cost}(P', \pi_1^{(1)}(X^*)) && (52)\\
&\le \frac{(1+\epsilon)^9}{1-\epsilon}\text{cost}(P, X^*), && (53)
\end{aligned}$$

where (47) is by Lemma 4.1, (48) is because (S, Δ, w) is an ϵ-coreset of P', (49) is by Lemma 4.2, (50) is because $\pi_1^{(2)}\circ\pi_1^{(1)}(X) = X'$, which is optimal in minimizing $\text{cost}((S', \Delta, w), \cdot)$, (51) is by Lemma 4.2, (52) is because (S, Δ, w) is an ϵ-coreset of P', and (53) is by Lemma 4.1.

For 2), note that by Theorem 3.2, the cardinality of the coreset constructed by FSS is $n' = O(k^3\log^2 k\epsilon^{-4}\log(1/\delta))$, which is independent of the dimension of the input dataset. Thus, the communication cost remains the same as that of Algorithm 2, which is $\tilde{O}(k^3/\epsilon^6)$.

For 3), note that the first JL projection $\pi_1^{(1)}$ takes $O(ndd')$ time, where $d' = O(\log n/\epsilon^2)$ by Lemma 4.1, and the second JL projection $\pi_1^{(2)}$ takes $O(n'd'd'')$ time, where n' is specified by Theorem 3.2 as above and $d'' = O(\epsilon^{-2}\log(n'k/\delta))$ by Lemma 4.2. Moreover, from the proof of Theorem 4.2, we know that applying FSS after a JL projection takes $O(\frac{n}{\epsilon^2}(\log^2 n/\epsilon^2 + k\log n/\epsilon^2 + k^2\log\frac{1}{\delta}))$ time. Thus, the total complexity at the data source is

$$\begin{aligned}
&O\left(ndd' + \frac{n}{\epsilon^2}\Big(\frac{\log^2 n}{\epsilon^2} + \frac{k\log n}{\epsilon^2} + k^2\log\frac{1}{\delta}\Big) + n'd'd''\right)\\
&= O\left(\frac{nd\log n}{\epsilon^2} + \frac{n\log^2 n}{\epsilon^4}\right) = \tilde{O}\left(\frac{nd}{\epsilon^2}\right). && (54)
\end{aligned}$$

□

A.7 Proof of Theorem 5.3

Proof. For 1), let $P := \bigcup_{i=1}^m P_i$, $\tilde{P} := \bigcup_{i=1}^m \tilde{P}_i$, and $\mathbf{S} := (S, 0, w)$ be the output of disSS. Let X^* be the optimal k-means centers of P. By Theorem 5.2, we know that with probability $\ge 1-\delta$,

$$\begin{aligned}
\text{cost}(\tilde{P}, X) \le \frac{1}{1-\epsilon}\text{cost}(\mathbf{S}, X) &\le \frac{1}{1-\epsilon}\text{cost}(\mathbf{S}, X^*)\\
&\le \frac{1+\epsilon}{1-\epsilon}\text{cost}(\tilde{P}, X^*), && (55)
\end{aligned}$$

where the second inequality is because X is optimal for $\mathbf{S}$. Moreover, by Theorem 5.1, we have

$$(1-\epsilon)\text{cost}(P, X) - \Delta \le \text{cost}(\tilde{P}, X), \qquad (56)$$

$$\text{cost}(\tilde{P}, X^*) \le (1+\epsilon)\text{cost}(P, X^*) - \Delta. \qquad (57)$$

Combining (55, 56, 57) yields

$$(1-\epsilon)\text{cost}(P,X) - \Delta \le \frac{1+\epsilon}{1-\epsilon}\cdot((1+\epsilon)\text{cost}(P,X^*) - \Delta) \le \frac{(1+\epsilon)^2}{1-\epsilon}\text{cost}(P,X^*) - \Delta, \quad (58)$$

which gives the desired approximation factor.

For 2), note that disPCA incurs a cost of $O(m \cdot (k/\epsilon^2) \cdot d)$ for transmitting $O(k/\epsilon^2)$ vectors in $\mathbb{R}^d$ from each of the m data sources, and disSS incurs a cost of $O\left(\frac{k}{\epsilon^2}\cdot(\epsilon^{-4}(\frac{k^2}{\epsilon^2}+\log\frac{1}{\delta}) + mk\log\frac{mk}{\delta})\right)$ for transmitting $O(\epsilon^{-4}(\frac{k^2}{\epsilon^2}+\log\frac{1}{\delta}) + mk\log\frac{mk}{\delta})$ vectors in $\mathbb{R}^{O(k/\epsilon^2)}$. For $d \gg m, k, 1/\epsilon$, and $1/\delta$, the total communication cost is dominated by the cost of disPCA.

For 3), as computing the local SVD at data source i takes $O(n_i d \cdot \min(n_i, d))$ time, the complexity of disPCA at the data sources is $O(nd \cdot \min(n, d))$. The complexity of disSS at data source i is dominated by the computation of bicriteria approximation of $\tilde{P}_i$, which takes $O(n_i t_2 k \log\frac{1}{\delta}) = O(nk^2\epsilon^{-2}\log\frac{1}{\delta})$ according to [42]. For $\min(n, d) \gg m, k, 1/\epsilon$, and $1/\delta$, the overall complexity is dominated by that of disPCA. □

A.8 Proof of Lemma 5.1

Proof. Let $\tilde{P}$ be the projection of P using the principal components computed by disPCA. Then by Theorem 5.1, there exists $\Delta \ge 0$ such that

$$(1-\epsilon)\text{cost}(P,X) \le \text{cost}(\tilde{P},X)+\Delta \le (1+\epsilon)\text{cost}(P,X). \quad (59)$$

Moreover, by Theorem 5.2, $\mathbf{S}$ is an ϵ-coreset of $\tilde{P}$ with probability at least $1-\delta$. Multiplying (59) by $1-\epsilon$, we have

$$(1-\epsilon)^2\text{cost}(P,X) \le (1-\epsilon)\text{cost}(\tilde{P},X) + (1-\epsilon)\Delta \quad (60)$$
$$\le \text{cost}(\mathbf{S},X) + \Delta, \quad (61)$$

where we can obtain (61) from (60) because $\mathbf{S}$ is an ϵ-coreset of $\tilde{P}$. Similarly, multiplying (59) by $1+\epsilon$, we have

$$(1+\epsilon)^2\text{cost}(P,X) \ge (1+\epsilon)\text{cost}(\tilde{P},X) + (1+\epsilon)\Delta \ge \text{cost}(\mathbf{S},X) + \Delta. \quad (62)$$

Combining (61, 62) yields the desired bound. □

A.9 Proof of Theorem 5.4

Proof. For 1), let $\mathbf{S}' := (\bigcup_{i=1}^m S'_i, \Delta, w)$, where $(\bigcup_{i=1}^m S'_i, 0, w)$ is the overall coreset constructed by line 3 of Algorithm 4, and Δ is a constant satisfying Lemma 5.1 for the input dataset $\{P'_i\}_{i=1}^m$ as in line 3 of Algorithm 4. Let $P := \bigcup_{i=1}^m P_i$, and X^* be the optimal k-means centers for P. Then with probability $\ge (1-\delta)^2$, we have

$$\text{cost}(P,X) \le (1+\epsilon)^2\text{cost}(\pi_1(P),X') \quad (63)$$
$$\le \frac{(1+\epsilon)^2}{(1-\epsilon)^2}\text{cost}(\mathbf{S}',X') \quad (64)$$
$$\le \frac{(1+\epsilon)^2}{(1-\epsilon)^2}\text{cost}(\mathbf{S}',\pi_1(X^*)) \quad (65)$$
$$\le \frac{(1+\epsilon)^4}{(1-\epsilon)^2}\text{cost}(\pi_1(P),\pi_1(X^*)) \quad (66)$$
$$\le \frac{(1+\epsilon)^6}{(1-\epsilon)^2}\text{cost}(P,X^*), \quad (67)$$

where (63) is by Lemma 4.1 (note that $\pi_1(X) = X'$), (64) is by Lemma 5.1 (note that $\text{cost}(\mathbf{S}',X') = \text{cost}((\bigcup_{i=1}^m S'_i, 0, w), X') + \Delta$), (65) is because X' is optimal in minimizing $\text{cost}(\mathbf{S}',\cdot)$, (66) is again by Lemma 5.1, and (67) is again by Lemma 4.1.

For 2), only line 3 incurs communication cost. By Theorem 5.3, we know that applying BKLW to a distributed dataset $\{P'_i\}_{i=1}^m$ with dimension d' incurs a cost of $O(mkd'/\epsilon^2)$, and by Lemma 4.1, we know that $d' = O(\log n/\epsilon^2)$, which yields the desired result.

For 3), the JL projection at each data source incurs a complexity of $O(ndd') = O(nd\log n/\epsilon^2)$. By Theorem 5.3, applying BKLW incurs a complexity of $O(nd'\cdot\min(n,d')) = O(n\log^2 n/\epsilon^4)$ at each data source. Together, the complexity is $O(\frac{nd}{\epsilon^2}\log n + \frac{n}{\epsilon^4}\log^2 n) = \tilde{O}(nd/\epsilon^4)$. □

A.10 Proof of Theorem 6.1

Proof. We only present the proof for Algorithm 3 with the incorporation of quantization, as the proofs for the other algorithms are similar. Consider a coreset (S, Δ, w) and a set of k-means centers X. If we quantize S into S_{QT} with a maximum quantization error of Δ_{QT}, then for each coreset point $q \in S$ and its quantized version $q' \in S_{QT}$, we have $\|q - q'\| \le \Delta_{QT}$. On the other hand, from [6], the k-means cost function is $2\Delta_D$-Lipschitz-continuous, which yields $|\text{cost}(q,X) - \text{cost}(q',X)| \le 2\Delta_D\Delta_{QT}$. Thus, the difference in the k-means cost between the original and the quantized coresets is bounded by

$$|\text{cost}((S,\Delta,w),X) - \text{cost}((S_{QT},\Delta,w),X)| \le 2\Delta_D\Delta_{QT}\sum_{q\in S} w(q), \quad (68)$$

as $\text{cost}((S,\Delta,w),X) = \sum_{q\in S} w(q)\text{cost}(q,X) + \Delta$.

Following the arguments in the proof of Theorem 4.4, we see that with probability $\geq (1-\delta)^3$:

$$
\begin{aligned}
&\text{cost}(P, X) \\
&\leq (1+\epsilon_1^{(1)})^2 \text{cost}(P', \pi_1^{(1)}(X)) && (69)\\
&\leq \frac{(1+\epsilon_1^{(1)})^2}{1-\epsilon_2} \text{cost}((S, \Delta, w), \pi_1^{(1)}(X)) && (70)\\
&\leq \frac{(1+\epsilon_1^{(1)})^2 (1+\epsilon_1^{(2)})^2}{1-\epsilon_2} \text{cost}((S', \Delta, w), \pi_1^{(2)} \circ \pi_1^{(1)}(X)) && (71)\\
&\leq \frac{(1+\epsilon_1^{(1)})^2 (1+\epsilon_1^{(2)})^2}{1-\epsilon_2} \cdot \\
&\quad (\text{cost}((S'_{QT}, \Delta, w), \pi_1^{(2)} \circ \pi_1^{(1)}(X)) + 2n\Delta_D \Delta_{QT}) && (72)\\
&\leq \frac{(1+\epsilon_1^{(1)})^2 (1+\epsilon_1^{(2)})^2}{1-\epsilon_2} \cdot \\
&\quad (\text{cost}((S'_{QT}, \Delta, w), \pi_1^{(2)} \circ \pi_1^{(1)}(X^*)) + 2n\Delta_D \Delta_{QT}) && (73)\\
&\leq \frac{(1+\epsilon_1^{(1)})^2 (1+\epsilon_1^{(2)})^2}{1-\epsilon_2} \cdot \\
&\quad (\text{cost}((S', \Delta, w), \pi_1^{(2)} \circ \pi_1^{(1)}(X^*)) + 4n\Delta_D \Delta_{QT}) && (74)\\
&\leq \frac{(1+\epsilon_1^{(1)})^2 (1+\epsilon_1^{(2)})^4}{1-\epsilon_2} \text{cost}((S, \Delta, w), \pi_1^{(1)}(X^*)) \\
&\quad + \frac{(1+\epsilon_1^{(1)})^2 (1+\epsilon_1^{(2)})^2}{1-\epsilon_2} 4n\Delta_D \Delta_{QT} && (75)\\
&\leq \frac{(1+\epsilon_1^{(1)})^2 (1+\epsilon_2)(1+\epsilon_1^{(2)})^4}{1-\epsilon_2} \text{cost}(P', \pi_1^{(1)}(X^*)) \\
&\quad + \frac{(1+\epsilon_1^{(1)})^2 (1+\epsilon_1^{(2)})^2}{1-\epsilon_2} 4n\Delta_D \Delta_{QT} && (76)\\
&\leq \frac{(1+\epsilon_1^{(1)})^4 (1+\epsilon_2)(1+\epsilon_1^{(2)})^4}{1-\epsilon_2} \text{cost}(P, X^*) \\
&\quad + \frac{(1+\epsilon_1^{(1)})^2 (1+\epsilon_1^{(2)})^2}{1-\epsilon_2} 4n\Delta_D \Delta_{QT}, && (77)
\end{aligned}
$$

where (72) and (74) are by (68) and the property that the coreset (S, Δ, w) constructed by sensitivity sampling satisfies $\sum_{q \in S} w(q) = n$ (the cardinality of P)[8]. □

8. While the sensitivity sampling procedure in [11] only guarantees that $\mathbb{E}[\sum_{q \in S} w(q)] = n$ (expectation over S), a variation of this procedure proposed in [4] guarantees $\sum_{q \in S} w(q) = n$ deterministically. FSS based on the sampling procedure in [4] still generates an ϵ-coreset (with probability $\geq 1-\delta$) with a constant cardinality (precisely, $O(\frac{k^2}{\epsilon^6} \log(\frac{1}{\delta}))$).