Supplementary Material

A Additional Content

A.1 Update Formula with Coincidence for $q = 1$

The $q = 1$ case is solved by [Vardi and Zhang, 2000] and interested readers are referred to it for a detailed proof. We directly give the $q = 1$ formula as follows:

$$\tilde{\mathbf{T}}(\mathbf{y}_{(p)}) = \frac{\sum_{i\neq k} \eta_i \|\mathbf{y}_{(p)} - \mathbf{x}_i\|^{-1}\mathbf{x}_i}{\sum_{i\neq k} \eta_i \|\mathbf{y}_{(p)} - \mathbf{x}_i\|^{-1}}, \tag{27}$$

$$\mathbf{y}_{(p+1)} = (1-\lambda)\tilde{\mathbf{T}}(\mathbf{y}_{(p)}) + \lambda \mathbf{y}_{(p)},$$

$$\lambda = \min\left\{1, \frac{\eta_k}{\|\nabla D_1(\mathbf{y}_{(p)})\|}\right\}. \tag{28}$$

This strategy ensures $\mathbf{y}_{(p+1)} \neq \mathbf{y}_{(p)} \Leftrightarrow C_1(\mathbf{y}_{(p+1)}) < C_1(\mathbf{y}_{(p)})$ and $\mathbf{y}_{(p+1)} = \mathbf{y}_{(p)} \Leftrightarrow \mathbf{y}_{(p)} = \mathbf{M}$.

A.2 Solving Algorithm

Note that the multiplicities are changed from $\{\eta_i\}_{i=1}^m$ back to $\{\xi_i\}_{i=1}^m$ to simplify the expressions.

Algorithm 1 q-th power Weiszfeld algorithm without singularity (qPWAWS)

Require: Given m distinct data points $\{\mathbf{x}_i\}_{i=1}^m$, the corresponding multiplicities $\{\xi_i\}_{i=1}^m$, the order of power q, the reducing factor ρ and the tolerance threshold Tol.
1. Initialize with a starting point $\mathbf{y}_{(0)}$.
while 1 **do**
if $\mathbf{y}_{(p)} \notin \{\mathbf{x}_i\}_{i=1}^m$ **then**
2. Compute $\mathbf{y}_{(p+1)} = \frac{\sum_{i=1}^m \xi_i \|\mathbf{y}_{(p)} - \mathbf{x}_i\|^{q-2}\mathbf{x}_i}{\sum_{i=1}^m \xi_i \|\mathbf{y}_{(p)} - \mathbf{x}_i\|^{q-2}}$.
if $\mathbf{y}_{(p+1)} = \mathbf{y}_{(p)}$ **then**
3. $\mathbf{M} = \mathbf{y}_{(p)}$. Break.
end if
4. $p \leftarrow p + 1$.
else
5. Suppose $\mathbf{y}_{(p)} = \mathbf{x}_k$,
compute $\nabla D_q(\mathbf{x}_k) = \sum_{i\neq k} q\xi_i \|\mathbf{x}_k - \mathbf{x}_i\|^{q-2}(\mathbf{x}_k - \mathbf{x}_i)$.
if $q = 1$ **then**
if $\|\nabla D_1(\mathbf{x}_k)\| \leqslant \xi_k$ **then**
6. $\mathbf{M} = \mathbf{x}_k$. Break.
else
7. Compute $\tilde{\mathbf{T}}(\mathbf{x}_k) = \frac{\sum_{i\neq k} \xi_i \|\mathbf{x}_k - \mathbf{x}_i\|^{-1}\mathbf{x}_i}{\sum_{i\neq k} \xi_i \|\mathbf{x}_k - \mathbf{x}_i\|^{-1}}$,
$\lambda = \frac{\xi_k}{\|\nabla D_1(\mathbf{x}_k)\|}, \mathbf{y}_{(p+1)} = (1-\lambda)\tilde{\mathbf{T}}(\mathbf{x}_k) + \lambda\mathbf{x}_k$.
8. $p \leftarrow p + 1$.
end if
else
if $\|\nabla D_q(\mathbf{x}_k)\| = 0$ **then**
9. $\mathbf{M} = \mathbf{x}_k$. Break.
else
10. Set $w = 0$,
$\lambda_w = \min\left\{\frac{1}{q}\xi_k^{-\frac{1}{q-1}} \|\nabla D_q(\mathbf{x}_k)\|^{\frac{2-q}{q-1}}, 1\right\}$.
while $C_q(\mathbf{x}_k - \lambda_w \nabla D_q(\mathbf{x}_k)) \geqslant C_q(\mathbf{x}_k)$ **do**
11. $\lambda_{w+1} = \rho\lambda_w$. $w \leftarrow w + 1$.
end while
12. $\mathbf{y}_{(p+1)} = \mathbf{x}_k - \lambda_w \nabla D_q(\mathbf{x}_k)$. $p \leftarrow p + 1$.
end if
end if
end if
if $\|\mathbf{y}_{(p+1)} - \mathbf{y}_{(p)}\| / \|\mathbf{y}_{(p)}\| \leqslant Tol$ **then**
13. $\mathbf{M} = \mathbf{y}_{(p+1)}$. Break.
end if
end while
Ensure: The minimum point $\mathbf{M}$.

B Proofs

B.1 Proof of Theorem 3

To prove this theorem, we need the following lemma:

Lemma 14 ([Weiszfeld, 1937; Cooper, 1968; Chen, 1984; Aftab *et al.*, 2015])**.** *If $a_i > 0$ and $b_i > 0$, $0 < q < n$ and $\sum_{i=1}^m a_i^{q-n} b_i^n < \sum_{i=1}^m a_i^q$, then $\sum_{i=1}^m b_i^q \leqslant \sum_{i=1}^m a_i^q$ and the equality holds only when $a_i = b_i, \forall i$.*

Proof. Consider the following function $g(t)$:

$$g(t) = \sum_{i=1}^{m} a_i^{q-t} b_i^t, 0 \leqslant t \leqslant n. \tag{29}$$

The second derivative of g with respect to t is:

$$g''(t) = \sum_{i=1}^{m} a_i^{q-t} b_i^t (\log a_i - \log b_i)^2. \tag{30}$$

Since all the $a_i, b_i > 0$, then $g''(t) > 0$ and $g(t)$ is a strictly convex function unless $a_i = b_i, \forall i$. If $g(t)$ is a strictly convex function, then $g(n) < g(0)$ implies $g(q) < g(0)$. Thus the lemma is proven. □

Lemma 14 reveals the relation between the q-th power ($1 \leqslant q < 2$) cost in (13) and the following weighted 2-nd power cost:

$$\tilde{C}_q(\mathbf{y}) = \sum_{i=1}^{m} \eta_i^q \|\mathbf{y}_{(p)} - \mathbf{x}_i\|^{q-2} \|\mathbf{y} - \mathbf{x}_i\|^2. \tag{31}$$

Proof. $\tilde{C}_q(\mathbf{y})$ in (31) is a strictly convex function on $\mathbf{y}$. By taking the gradient of $\tilde{C}_q(\mathbf{y})$ and setting it to zero, it yields:

$$\nabla \tilde{C}_q(\mathbf{y}) = \sum_{i=1}^{m} 2\eta_i^q \|\mathbf{y}_{(p)} - \mathbf{x}_i\|^{q-2} (\mathbf{y} - \mathbf{x}_i) = \mathbf{0}. \tag{32}$$

Hence $\mathbf{T}_1(\mathbf{y}_{(p)})$ is the minimizer of $\tilde{C}_q(\mathbf{y})$. It yields $\tilde{C}_q(\mathbf{T}_1(\mathbf{y}_{(p)})) \leqslant \tilde{C}_q(\mathbf{y}_{(p)}) = C_q(\mathbf{y}_{(p)})$ with equality holds only when $\mathbf{T}_1(\mathbf{y}_{(p)}) = \mathbf{y}_{(p)}$.

If $\tilde{C}_q(\mathbf{T}_1(\mathbf{y}_{(p)})) < C_q(\mathbf{y}_{(p)})$, it means:

$$\begin{aligned} \sum_{i=1}^{m} \eta_i^q \|\mathbf{y}_{(p)} - \mathbf{x}_i\|^{q-2} \|\mathbf{T}_1(\mathbf{y}_{(p)}) - \mathbf{x}_i\|^2 &< \sum_{i=1}^{m} \eta_i^q \|\mathbf{y}_{(p)} - \mathbf{x}_i\|^q, \\ \sum_{i=1}^{m} \|\eta_i(\mathbf{y}_{(p)}-\mathbf{x}_i)\|^{q-2} \|\eta_i(\mathbf{T}_1(\mathbf{y}_{(p)})-\mathbf{x}_i)\|^2 &< \sum_{i=1}^{m} \|\eta_i(\mathbf{y}_{(p)}-\mathbf{x}_i)\|^q. \end{aligned} \tag{33}$$

By setting $a_i = \|\eta_i(\mathbf{y}_{(p)}-\mathbf{x}_i)\|$, $b_i = \|\eta_i(\mathbf{T}_1(\mathbf{y}_{(p)})-\mathbf{x}_i)\|$ for all i and using Lemma 14 ($n = 2$), it leads to:

$$C_q(\mathbf{T}_1(\mathbf{y}_{(p)})) = \sum_{i=1}^{m} \|\eta_i(\mathbf{T}_1(\mathbf{y}_{(p)})-\mathbf{x}_i)\|^q < \sum_{i=1}^{m} \|\eta_i(\mathbf{y}_{(p)}-\mathbf{x}_i)\|^q = C_q(\mathbf{y}_{(p)}). \tag{34}$$

It proves Theorem 3. □

B.2 Proof of Corollary 4

Proof. With $\mathbf{y}_{(p)} \notin \{\mathbf{x}_i\}_{i=1}^m$ and (14), the following equivalence holds:

$$\mathbf{T}_1(\mathbf{y}_{(p)}) = \mathbf{y}_{(p)} \quad \Longleftrightarrow \quad \mathbf{0} = \sum_{i=1}^{m} \eta_i^q \|\mathbf{y}_{(p)} - \mathbf{x}_i\|^{q-2} (\mathbf{y}_{(p)} - \mathbf{x}_i) = \frac{1}{q} \nabla C_q(\mathbf{y}_{(p)}). \tag{35}$$

Since $C_q(\mathbf{y})$ is strictly convex, $\nabla C_q(\mathbf{y}_{(p)}) = \mathbf{0} \Leftrightarrow \mathbf{y}_{(p)} = \mathbf{M}$. □

B.3 Proof of Theorem 6

Proof. Let $\mathbf{x}_k + \lambda\mathbf{z}$ ($\lambda > 0, \|\mathbf{z}\| = 1$) be a point displaced from $\mathbf{x}_k$ towards an arbitrary direction. Then the gradient of $C_q(\mathbf{x}_k + \lambda\mathbf{z})$ with respect to λ is:

$$\frac{\mathrm{d}C_q(\mathbf{x}_k + \lambda\mathbf{z})}{\mathrm{d}\lambda} = \sum_{i\neq k} q\eta_i^q \|\mathbf{x}_k + \lambda\mathbf{z} - \mathbf{x}_i\|^{q-2}(\mathbf{x}_k + \lambda\mathbf{z} - \mathbf{x}_i)^\top \mathbf{z} + q\eta_k^q \lambda^{q-1}. \tag{36}$$

The limit of $\frac{\mathrm{d}}{\mathrm{d}\lambda}C_q(\mathbf{x}_k + \lambda\mathbf{z})$ when $\lambda \to 0$ is:

$$\frac{\mathrm{d}C_1(\mathbf{x}_k + \lambda\mathbf{z})}{\mathrm{d}\lambda}\Big|_{\lambda=0} = \sum_{i\neq k} \eta_i \|\mathbf{x}_k - \mathbf{x}_i\|^{-1}(\mathbf{x}_k - \mathbf{x}_i)^\top \mathbf{z} + \eta_k, \quad q = 1. \tag{37a}$$

$$\frac{\mathrm{d}C_q(\mathbf{x}_k + \lambda\mathbf{z})}{\mathrm{d}\lambda}\Big|_{\lambda=0} = \sum_{i\neq k} q\eta_i^q \|\mathbf{x}_k - \mathbf{x}_i\|^{q-2}(\mathbf{x}_k - \mathbf{x}_i)^\top \mathbf{z}, \quad 1 < q < 2. \tag{37b}$$

From Definition 5, (37a) and (37b) can be formulated as:

$$\frac{\mathrm{d}C_1(\mathbf{x}_k + \lambda\mathbf{z})}{\mathrm{d}\lambda}\Big|_{\lambda=0} = \nabla D_1(\mathbf{x}_k)^\top \mathbf{z} + \eta_k, \quad q = 1. \tag{38a}$$

$$\frac{\mathrm{d}C_q(\mathbf{x}_k + \lambda\mathbf{z})}{\mathrm{d}\lambda}\Big|_{\lambda=0} = \nabla D_q(\mathbf{x}_k)^\top \mathbf{z}, \quad 1 < q < 2. \tag{38b}$$

Thus the multiplicity η_k affects the gradient only when $q = 1$. By setting $\mathbf{z} = -\dfrac{\nabla D_q(\mathbf{x}_k)}{\|\nabla D_q(\mathbf{x}_k)\|}$ in (38a) and (38b), we have:

$$\min_{\mathbf{z}} \frac{\mathrm{d}C_1(\mathbf{x}_k + \lambda\mathbf{z})}{\mathrm{d}\lambda}\Big|_{\lambda=0} = -\|\nabla D_1(\mathbf{x}_k)\| + \eta_k, \quad q = 1. \tag{39a}$$

$$\min_{\mathbf{z}} \frac{\mathrm{d}C_q(\mathbf{x}_k + \lambda\mathbf{z})}{\mathrm{d}\lambda}\Big|_{\lambda=0} = -\|\nabla D_q(\mathbf{x}_k)\|, \quad 1 < q < 2. \tag{39b}$$

$$C_q(\mathbf{x}_k) \text{ is the minimum} \quad \Longleftrightarrow \quad \min_{\mathbf{z}} \frac{\mathrm{d}C_q(\mathbf{x}_k + \lambda\mathbf{z})}{\mathrm{d}\lambda}\Big|_{\lambda=0} \geqslant 0,\ 1 \leqslant q < 2. \tag{39c}$$

Combining (39a), (39b) and (39c), one can find that the subgradient sets in (17) are equivalent to Definition 1. Thus Theorem 6 is proven. □

B.4 Proof of Theorem 7

Proof. This theorem indicates that a sufficiently small displacement towards the negative de-singularity subgradient $-\nabla D_q(\mathbf{x}_k)$ can reduce the q-th power cost. $C_q(\mathbf{x}_k - \lambda\nabla D_q(\mathbf{x}_k))$ is continuous on $\lambda > 0$. It consists of two parts: the nonsingular part $D_q(\mathbf{x}_k - \lambda\nabla D_q(\mathbf{x}_k))$ and the singular part $\eta_k^q\lambda^q\|\nabla D_q(\mathbf{x}_k)\|^q$. From the Taylor series expansion of $D_q(\mathbf{x}_k - \lambda\nabla D_q(\mathbf{x}_k))$,

$$\begin{aligned}
&C_q(\mathbf{x}_k - \lambda\nabla D_q(\mathbf{x}_k)) \\
=&D_q(\mathbf{x}_k - \lambda\nabla D_q(\mathbf{x}_k)) + \eta_k^q\lambda^q\|\nabla D_q(\mathbf{x}_k)\|^q \\
=&D_q(\mathbf{x}_k) - \lambda\|\nabla D_q(\mathbf{x}_k)\|^2 + \frac{\lambda^2}{2}\nabla D_q(\mathbf{x}_k)^\top H(\mathbf{x}_k)\nabla D_q(\mathbf{x}_k) + o(\lambda^2) + \eta_k^q\lambda^q\|\nabla D_q(\mathbf{x}_k)\|^q,
\end{aligned} \tag{40}$$

where $H(\mathbf{x}_k)$ is the Hessian of $D_q(\mathbf{y})$ at $\mathbf{x}_k$. Besides, it is easy to find that $D_q(\mathbf{x}_k) = C_q(\mathbf{x}_k)$. Then (40) can be rearranged to:

$$C_q(\mathbf{x}_k - \lambda\nabla D_q(\mathbf{x}_k)) - C_q(\mathbf{x}_k) = -\lambda\|\nabla D_q(\mathbf{x}_k)\|^2 + \frac{\lambda^2}{2}G(\mathbf{x}_k) + o(\lambda^2) + \eta_k^q\lambda^q\|\nabla D_q(\mathbf{x}_k)\|^q, \tag{41}$$

where $G(\mathbf{x}_k) \triangleq \nabla D_q(\mathbf{x}_k)^\top H(\mathbf{x}_k)\nabla D_q(\mathbf{x}_k)$. Therefore, we need to find a λ such that the right side of (41) is negative.

When $\lambda \to 0$, the negative term $-\lambda\|\nabla D_q(\mathbf{x}_k)\|^2$ dominates the other terms on the right side, thus it is possible to make the right side negative. To specify, a λ should be found to satisfy the following inequality:

$$-\lambda\|\nabla D_q(\mathbf{x}_k)\|^2 + \frac{\lambda^2}{2}G(\mathbf{x}_k) + o(\lambda^2) + \eta_k^q\lambda^q\|\nabla D_q(\mathbf{x}_k)\|^q < 0. \tag{42}$$

Dividing both sides of (42) by λ yields:

$$\begin{aligned}
&-\|\nabla D_q(\mathbf{x}_k)\|^2 + \frac{\lambda}{2}G(\mathbf{x}_k) + o(\lambda) + \eta_k^q\lambda^{q-1}\|\nabla D_q(\mathbf{x}_k)\|^q < 0, \\
&\frac{\lambda}{2}G(\mathbf{x}_k) + o(\lambda) + \eta_k^q\lambda^{q-1}\|\nabla D_q(\mathbf{x}_k)\|^q < \|\nabla D_q(\mathbf{x}_k)\|^2.
\end{aligned} \tag{43}$$

Since $1 < q < 2$, the left side of (43) approaches zero when $\lambda \to 0$, while $\|\nabla D_q(\mathbf{x}_k)\|^2 > 0$. Therefore, there exists a $\lambda_* > 0$ such that for any $0 < \lambda \leqslant \lambda_*$, (43) holds. From (43) back to (41), the theorem is proven. □

B.5 Proof of Lemma 8

Proof. From Corollary 4, Theorem 3, Theorem 6 and Theorem 7, qPWAWS has the following decreasing property:

$$C_q(\mathbf{y}_{(0)}) > C_q(\mathbf{y}_{(1)}) > \cdots > C_q(\mathbf{y}_{(p)}) > \cdots > C_q(\mathbf{M}), \tag{44}$$

unless some $\mathbf{y}_{(p)}$ hits $\mathbf{M}$. In particular, if $\mathbf{y}_{(p)} = \mathbf{x}_k$ but $\mathbf{x}_k \neq \mathbf{M}$, then $C_q(\mathbf{y}_{(p)}) > C_q(\mathbf{y}_{(p+1)})$ and the subsequent iterates will never get back to $\mathbf{x}_k$, otherwise the decreasing property will be violated. Hence the sequence of iterates visits each $\mathbf{x}_k \neq \mathbf{M}$ at most once and will not get stuck.

From (20), if $\mathbf{y}_{(p)} \notin \{\mathbf{x}_i\}_{i=1}^m$, $\mathbf{T}_1(\mathbf{y}_{(p)})$ is a weighted sum of the data points $\{\mathbf{x}_i\}_{i=1}^m$ with positive weights that sum to one. Hence $\mathbf{y}_{(p+1)}=\mathbf{T}_1(\mathbf{y}_{(p)})$ lies in the convex hull of $\{\mathbf{x}_i\}_{i=1}^m$. Since $\{\mathbf{y}_{(p)}\}$ visits each $\mathbf{x}_k \neq \mathbf{M}$ at most once, $\mathbf{T}_2(\mathbf{y}_{(p)})$ is invoked at most finite times. Because $\mathbf{T}_2(\mathbf{y}_{(p)})$ cannot ensure that $\mathbf{y}_{(p+1)}$ lies in the convex hull, there are at most a finite set of iterates that do not lie in the convex hull.

Last, if $\mathbf{M} \in \{\mathbf{x}_i\}_{i=1}^m$, then $\mathbf{M}$ is trivially in the convex hull. If $\mathbf{M} \notin \{\mathbf{x}_i\}_{i=1}^m$, Corollary 4 and (35) indicate that $\mathbf{M}$ lies in the convex hull. □

B.6 Proof of Lemma 9

Proof. First, taking a difference between both sides of (21) leads to

$$\mathbf{T}_1(\mathbf{y}) - \mathbf{x}_k = \frac{\sum_{i\neq k} \eta_i^q \|\mathbf{y} - \mathbf{x}_i\|^{q-2}(\mathbf{x}_i - \mathbf{x}_k)}{\sum_{i=1}^m \eta_i^q \|\mathbf{y} - \mathbf{x}_i\|^{q-2}} \tag{45}$$

$$= \frac{\|\mathbf{y} - \mathbf{x}_k\|^{2-q} \cdot (\sum_{i\neq k} \eta_i^q \|\mathbf{y} - \mathbf{x}_i\|^{q-2}(\mathbf{x}_i - \mathbf{x}_k))}{\eta_k^q + \|\mathbf{y} - \mathbf{x}_k\|^{2-q} \cdot (\sum_{i\neq k} \eta_i^q \|\mathbf{y} - \mathbf{x}_i\|^{q-2})}. \tag{46}$$

Then the limit of its L_2-norm is

$$\lim_{\mathbf{y}\to\mathbf{x}_k} \|\mathbf{T}_1(\mathbf{y}) - \mathbf{x}_k\| = \frac{0 \cdot \| \sum_{i\neq k} \eta_i^q \|\mathbf{x}_k - \mathbf{x}_i\|^{q-2}(\mathbf{x}_i - \mathbf{x}_k)\|}{\eta_k^q + 0 \cdot (\sum_{i\neq k} \eta_i^q \|\mathbf{x}_k - \mathbf{x}_i\|^{q-2})} = 0. \tag{47}$$

Since $\eta_k^q \neq 0$, the above limit is well-defined and equals 0. It indicates that $\mathbf{T}_1(\mathbf{y}) \to \mathbf{x}_k$ when $\mathbf{y} \to \mathbf{x}_k$. □

B.7 Proof of Lemma 10

Proof. From (6), if $\mathbf{x}_k \neq \mathbf{M}$, then $\|\nabla D_q(\mathbf{x}_k)\| > 0$. This is the key condition for $\mathbf{T}_1$ to drive $\mathbf{y}$ away. For any sufficiently small $0 < \epsilon < 1$, since $\|\nabla D_q(\mathbf{y})\|$ is continuous around $\mathbf{x}_k$, there exists $\delta_1 > 0$ such that

$$\mathbf{y} \in B(\mathbf{x}_k, \delta_1) \Longrightarrow \|\nabla D_q(\mathbf{y})\| > \|\nabla D_q(\mathbf{x}_k)\| - \epsilon > 0. \tag{48}$$

Second, when $\mathbf{y} \to \mathbf{x}_k$, the weight of $\mathbf{x}_k$ in (14) will approach 1. In other words, there exists $\delta_2 > 0$ such that

$$\mathbf{y} \in B(\mathbf{x}_k, \delta_2) \Longrightarrow 1 - \epsilon < \frac{\eta_k^q \|\mathbf{y} - \mathbf{x}_k\|^{q-2}}{\sum_{i=1}^m \eta_i^q \|\mathbf{y} - \mathbf{x}_i\|^{q-2}} < 1. \tag{49}$$

Third, to handle some remainders, define $\delta_3 > 0$ as follows:

$$\delta_3 = \left(\frac{(1-\epsilon)(\|\nabla D_q(\mathbf{x}_k)\| - \epsilon)}{q\eta_k^q (1 + 2\epsilon)} \right)^{1/(q-1)}, \quad 1 < q < 2. \tag{50}$$

$$\mathbf{y} \in B(\mathbf{x}_k, \delta_3) \Longrightarrow \frac{(1-\epsilon)(\|\nabla D_q(\mathbf{x}_k)\| - \epsilon)}{q\eta_k^q \|\mathbf{y} - \mathbf{x}_k\|^{q-1}} > 1 + 2\epsilon. \tag{51}$$

When $0 < \epsilon < 1$ is sufficiently small, $\delta_3 > 0$ is well defined.

Let $\delta_0 = \min\{\delta_1, \delta_2, \delta_3\}$ and $\mathbf{y} \in B(\mathbf{x}_k, \delta_0)$, then:

$$\mathbf{T}_1(\mathbf{y}) - \mathbf{x}_k = \frac{\sum_{i=1}^m \eta_i^q \|\mathbf{y} - \mathbf{x}_i\|^{q-2}(\mathbf{x}_i - \mathbf{y})}{\sum_{i=1}^m \eta_i^q \|\mathbf{y} - \mathbf{x}_i\|^{q-2}} + \mathbf{y} - \mathbf{x}_k$$

$$= \frac{-\nabla D_q(\mathbf{y})/q}{\sum_{i=1}^m \eta_i^q \|\mathbf{y} - \mathbf{x}_i\|^{q-2}} + \left(\frac{\eta_k^q \|\mathbf{y} - \mathbf{x}_k\|^{q-2}}{\sum_{i=1}^m \eta_i^q \|\mathbf{y} - \mathbf{x}_i\|^{q-2}} - 1 \right)(\mathbf{x}_k - \mathbf{y}). \tag{52}$$

Therefore

$$\|\mathbf{T}_1(\mathbf{y}) - \mathbf{x}_k\| > \frac{\|\nabla D_q(\mathbf{y})/q\|}{\sum_{i=1}^m \eta_i^q \|\mathbf{y} - \mathbf{x}_i\|^{q-2}} - \epsilon \|\mathbf{x}_k - \mathbf{y}\|$$

$$
\begin{aligned}
&> \frac{(1-\epsilon)\|\nabla D_q(\mathbf{y})/q\|}{\eta_k^q \|\mathbf{y}-\mathbf{x}_k\|^{q-2}} - \epsilon\|\mathbf{x}_k - \mathbf{y}\| \\
&> \frac{(1-\epsilon)(\|\nabla D_q(\mathbf{x}_k)\| - \epsilon)}{q\eta_k^q \|\mathbf{y}-\mathbf{x}_k\|^{q-2}} - \epsilon\|\mathbf{x}_k - \mathbf{y}\| \\
&> (1+2\epsilon)\|\mathbf{x}_k - \mathbf{y}\| - \epsilon\|\mathbf{x}_k - \mathbf{y}\| \\
&= (1+\epsilon)\|\mathbf{x}_k - \mathbf{y}\|,
\end{aligned} \tag{53}
$$

where the first inequality is based on the triangle inequality and (49); The second inequality is based on the left inequality of (49); The third and the fourth inequalities are based on (48) and (51), respectively.

Therefore, $\|\mathbf{T}_1(\mathbf{y}) - \mathbf{x}_k\| > (1+\epsilon)\|\mathbf{y} - \mathbf{x}_k\|$. If $\mathbf{T}_1(\mathbf{y}) \in B(\mathbf{x}_k, \delta_0)$, then $\|\mathbf{T}_1^2(\mathbf{y}) - \mathbf{x}_k\| > (1+\epsilon)\|\mathbf{T}_1(\mathbf{y}) - \mathbf{x}_k\| > (1+\epsilon)^2\|\mathbf{y} - \mathbf{x}_k\|$. As long as the current iterate lies in $B(\mathbf{x}_k, \delta_0)$, $\mathbf{T}_1$ will keep on driving it out of $B(\mathbf{x}_k, \delta_0)$. Thus there exists some s such that $\mathbf{T}_1^{s-1}(\mathbf{y}) \in B(\mathbf{x}_k, \delta_0)$ and $\mathbf{T}_1^s(\mathbf{y}) \notin B(\mathbf{x}_k, \delta_0)$ (see Figure 5). $\square$

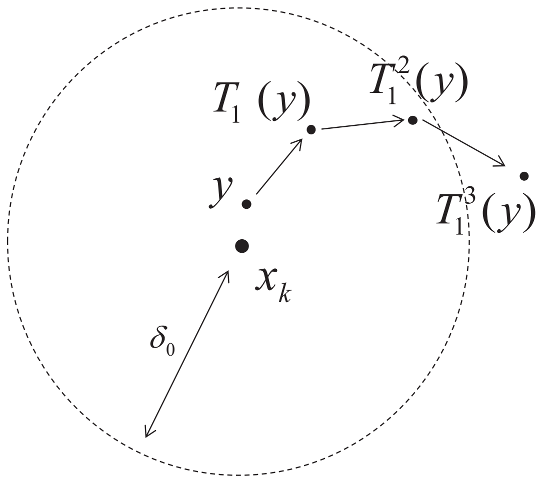

Figure 5: Once $\mathbf{y}$ gets into $B(\mathbf{x}_k, \delta_0)$ where $\mathbf{x}_k \neq \mathbf{M}$, $\mathbf{T}_1$ will eventually drive it out of $B(\mathbf{x}_k, \delta_0)$.

B.8 Proof of Theorem 11

Proof. We can assume that $\mathbf{y}_{(p)}$ differs from $\mathbf{M}$ for all p. From Lemma 8, since at most a finite set of iterates do not lie in the convex hull of $\{\mathbf{x}_i\}_{i=1}^m$, the whole sequence $\{\mathbf{y}_{(p)}\}$ is a compact set in $\mathbb{R}^d$. Moreover, by omitting at most a finite number of iterates, we can assume that $\{\mathbf{y}_{(p)}\} \bigcap \{\mathbf{x}_i\}_{i=1}^m = \varnothing$. By the Bolzano-Weierstrass Theorem, there exists a subsequence $\{\mathbf{y}_{(p_v)}\}$ such that $\lim_{v\to\infty} \mathbf{y}_{(p_v)} = \mathbf{y}_*$ for some $\mathbf{y}_* \in \mathbb{R}^d$. Since the extended operator $\mathbf{T}_1$ is continuous,

$$\lim_{v\to\infty} \mathbf{T}_1(\mathbf{y}_{(p_v)}) = \mathbf{T}_1(\mathbf{y}_*). \tag{54}$$

According to the decreasing property of qPWAWS (44), the sequence $C_q(\mathbf{y}_{(p)})$ is bounded below and decreasing, thus it has a limit and any subsequence of $C_q(\mathbf{y}_{(p)})$ should have the same limit. In particular, $C_q(\mathbf{y}_{(p_v)})$ and $C_q(\mathbf{T}_1(\mathbf{y}_{(p_v)}))$ are two subsequences of $C_q(\mathbf{y}_{(p)})$. Hence

$$\lim_{v\to\infty} C_q(\mathbf{T}_1(\mathbf{y}_{(p_v)})) = \lim_{v\to\infty} C_q(\mathbf{y}_{(p_v)}). \tag{55}$$

Since C_q is continuous, (54) and (55) indicate

$$C_q(\mathbf{T}_1(\mathbf{y}_*)) = C_q(\mathbf{y}_*). \tag{56}$$

If $\mathbf{y}_* \notin \{\mathbf{x}_i\}_{i=1}^m$, then Theorem 3 and (56) indicate $\mathbf{y}_* = \mathbf{T}_1(\mathbf{y}_*)$. By Corollary 4, $\mathbf{y}_* = \mathbf{M}$. If $\mathbf{y}_* \in \{\mathbf{x}_i\}_{i=1}^m$, then (56) trivially holds from (22). To summarize, $\mathbf{y}_* \in \{\mathbf{x}_i\}_{i=1}^m \bigcup \{\mathbf{M}\}$ and only the points in the finite set $\{\mathbf{x}_i\}_{i=1}^m \bigcup \{\mathbf{M}\}$ satisfy (56) and constitute the fixed points of $\mathbf{T}_1$.

The next step is to prove that if $\mathbf{y}_* = \mathbf{x}_k$ for some k, then $\mathbf{x}_k = \mathbf{M}$. If not, we invoke Lemma 10 to induce a contradiction. Since $\lim_{v\to\infty} \mathbf{y}_{(p_v)} = \mathbf{x}_k$, once $\mathbf{y}_{(p_v)}$ gets into $B(\mathbf{x}_k, \delta_0)$, it will be driven out by $\mathbf{T}_1$. Thus for each $\mathbf{y}_{(p_v)} \in B(\mathbf{x}_k, \delta_0)$, there exists a $\mathbf{y}_{(p_u)} \in B(\mathbf{x}_k, \delta_0)$ and a $\mathbf{T}_1(\mathbf{y}_{(p_u)}) \notin B(\mathbf{x}_k, \delta_0)$. In other words, $\mathbf{y}_{(p_u)}$ is the iterate that is going to be driven out of $B(\mathbf{x}_k, \delta_0)$. Since $\mathbf{y}_{(p_v)}$ is an infinite sequence converging to $\mathbf{x}_k$, $\mathbf{y}_{(p_u)}$ and $\mathbf{T}_1(\mathbf{y}_{(p_u)})$ are also infinite sequences. By the

Bolzano-Weierstrass Theorem, $\mathbf{y}_{(p_u)}$ has a subsequence that converges to some $\mathbf{y}_{*1}$. We still denote this subsequence by $\mathbf{y}_{(p_u)}$. Then $\mathbf{T}_1(\mathbf{y}_{(p_u)})$ also has a subsequence that converges to some $\mathbf{y}_{*2}$ and the subsequence can still be denoted by $\mathbf{T}_1(\mathbf{y}_{(p_u)})$. Note that $\mathbf{y}_{(p_u)}$ and $\mathbf{T}_1(\mathbf{y}_{(p_u)})$ are not necessarily subsequences of $\mathbf{y}_{(p_v)}$. Then

$$\lim_{u\to\infty} \mathbf{y}_{(p_u)} = \mathbf{y}_{*1}, \quad \lim_{u\to\infty} \mathbf{T}_1(\mathbf{y}_{(p_u)}) = \mathbf{y}_{*2}. \tag{57}$$

$$\|\mathbf{y}_{*1} - \mathbf{x}_k\| \leqslant \delta_0, \quad \|\mathbf{y}_{*2} - \mathbf{x}_k\| \geqslant \delta_0. \tag{58}$$

Similar to the demonstrations of (54),(55) and (56), the accumulation point $\mathbf{y}_{*1}$ is also a fixed point of $\mathbf{T}_1$:

$$\lim_{u\to\infty} \mathbf{T}_1(\mathbf{y}_{(p_u)}) = \mathbf{T}_1(\lim_{u\to\infty} \mathbf{y}_{(p_u)}) = \mathbf{T}_1(\mathbf{y}_{*1}) = \mathbf{y}_{*1}. \tag{59}$$

From (57), (58) and (59),

$$\mathbf{y}_{*1} = \mathbf{y}_{*2}, \quad \|\mathbf{y}_{*1} - \mathbf{x}_k\| = \delta_0. \tag{60}$$

The process of inducing $\mathbf{y}_{*1} = \mathbf{y}_{*2}$ can be shown as Figure 6.

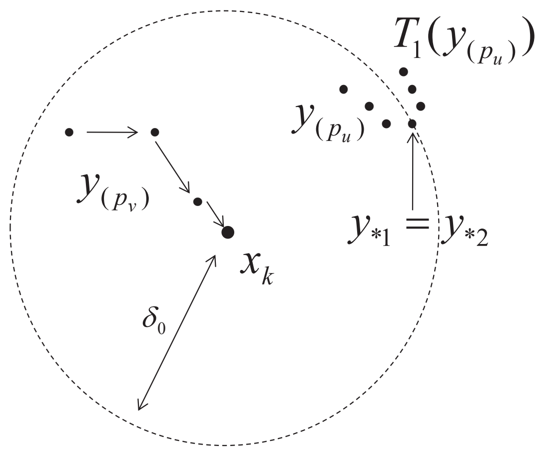

Figure 6: The process of inducing $\mathbf{y}_{*1} = \mathbf{y}_{*2}$. $\mathbf{y}_{(p_v)}$ is a subsequence that converges to $\mathbf{x}_k \neq \mathbf{M}$. By Lemma 10, each $\mathbf{y}_{(p_v)}$ induces a $\mathbf{y}_{(p_u)} \in B(\mathbf{x}_k, \delta_0)$ and a $\mathbf{T}_1(\mathbf{y}_{(p_u)}) \notin B(\mathbf{x}_k, \delta_0)$. Then the subsequences $\mathbf{y}_{(p_u)}$ and $\mathbf{T}_1(\mathbf{y}_{(p_u)})$ have the same accumulation point $\mathbf{y}_{*1} = \mathbf{y}_{*2}$ at the boundary of $B(\mathbf{x}_k, \delta_0)$.

For any $0 < \delta < \delta_0$, we can apply the same method to obtain a distinct fixed point $\mathbf{y}_{*\delta}$ such that $\mathbf{T}_1(\mathbf{y}_{*\delta}) = \mathbf{y}_{*\delta}$ and $\|\mathbf{y}_{*\delta} - \mathbf{x}_k\| = \delta$. Then there are infinite fixed points $\{\mathbf{y}_{*\delta}\}$, which is contradictory to the finite set of fixed points $\{\mathbf{x}_i\}_{i=1}^m \bigcup \{\mathbf{M}\}$. Therefore, $\mathbf{y}_* = \mathbf{x}_k = \mathbf{M}$.

The last step is to prove that the whole sequence $\{\mathbf{y}_{(p)}\}$ converges to $\mathbf{y}_*$. From the above illustrations, the accumulation point $\mathbf{y}_* = \mathbf{M}$. If there is another accumulation point $\tilde{\mathbf{y}} \neq \mathbf{y}_*$, then $\tilde{\mathbf{y}} \in \{\mathbf{x}_i\}_{i=1}^m$. Without loss of generality, suppose $\tilde{\mathbf{y}} = \mathbf{x}_k \neq \mathbf{M}$, then the above method can be repeated to induce an infinite set of fixed points $\{\mathbf{y}_{*\delta}\}$, which leads to a contradiction. Hence, there is only one accumulation point $\mathbf{y}_* = \mathbf{M}$ for the whole sequence $\{\mathbf{y}_{(p)}\}$. Thus $\{\mathbf{y}_{(p)}\}$ and any subsequences converge to $\mathbf{y}_* = \mathbf{M}$. □

B.9 Proof of Lemma 12

Proof. It is straightforward to check from (16) that $\nabla D_q(\mathbf{y})$ is analytic in some neighborhood $B(\mathbf{x}_k, \delta)$ of $\mathbf{x}_k$ such that $B(\mathbf{x}_k, \delta) \bigcap \{\mathbf{x}_i\}_{i\neq k} = \varnothing$, since the singular component has been excluded from $\nabla D_q(\mathbf{y})$. Furthermore, $\|\nabla D_q(\mathbf{y})\|^2 = \nabla D_q(\mathbf{y})^\top \nabla D_q(\mathbf{y})$ is also analytic in this neighborhood $B(\mathbf{x}_k, \delta)$. Thus we can adopt the second-order Taylor series expansion of $\|\nabla D_q(\mathbf{y})\|^2$ for $\mathbf{y}_{(p)} \in B(\mathbf{x}_k, \delta)$ at $\mathbf{x}_k$:

$$\|\nabla D_q(\mathbf{y}_{(p)})\|^2 = \|\nabla D_q(\mathbf{x}_k)\|^2 + 2\nabla D_q(\mathbf{x}_k)^\top H(\mathbf{x}_k)(\mathbf{y}_{(p)} - \mathbf{x}_k) + \frac{1}{2}(\mathbf{y}_{(p)} - \mathbf{x}_k)^\top J(\mathbf{x}_k)(\mathbf{y}_{(p)} - \mathbf{x}_k) + o(\|\mathbf{y}_{(p)} - \mathbf{x}_k\|^2), \tag{61}$$

where $H(\mathbf{x}_k)$ and $J(\mathbf{x}_k)$ are the Hessians of $D_q(\mathbf{y})$ and $\|\nabla D_q(\mathbf{y})\|^2$ at $\mathbf{x}_k$, respectively.

Theorem 6 indicates that $\nabla D_q(\mathbf{x}_k) = \mathbf{0}$ when $1 < q < 2$, thus (61) can be further simplified as

$$\|\nabla D_q(\mathbf{y}_{(p)})\|^2 = \frac{1}{2}(\mathbf{y}_{(p)} - \mathbf{x}_k)^\top J(\mathbf{x}_k)(\mathbf{y}_{(p)} - \mathbf{x}_k) + o(\|\mathbf{y}_{(p)} - \mathbf{x}_k\|^2) \leqslant \frac{\vartheta_J}{2}\|\mathbf{y}_{(p)} - \mathbf{x}_k\|^2 + o(\|\mathbf{y}_{(p)} - \mathbf{x}_k\|^2), \tag{62}$$

where ϑ_J denotes the largest eigenvalue of $J(\mathbf{x}_k)$. Since $\|\nabla D_q(\mathbf{y}_{(p)})\|^2>0$ when $\mathbf{y}_{(p)}\neq\mathbf{x}_k$, (62) implies that

$$\frac{\vartheta_J}{2}\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^2+o(\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^2)\geqslant\|\nabla D_q(\mathbf{y}_{(p)})\|^2>0\quad\Longrightarrow\quad\frac{\vartheta_J}{2}+o(1)>0\quad\Longrightarrow\quad\vartheta_J\geqslant 0. \tag{63}$$

It means that the inequality in (62) really holds without contradiction.

Next, dividing the leftmost side and the rightmost side of (62) by $\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^2$ leads to

$$\frac{\|\nabla D_q(\mathbf{y}_{(p)})\|^2}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^2}\leqslant\frac{\vartheta_J}{2}+o(1)\quad\Longrightarrow\quad\lim_{\mathbf{y}_{(p)}\to\mathbf{x}_k}\frac{\|\nabla D_q(\mathbf{y}_{(p)})\|^2}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^2}\leqslant\frac{\vartheta_J}{2}. \tag{64}$$

Since the square root operator $\sqrt{\cdot}$ is continuous in the interval $(0,+\infty)$ and right continuous at 0, we can take $\sqrt{\cdot}$ inside the limit of (64) and get

$$\lim_{\mathbf{y}_{(p)}\to\mathbf{x}_k}\frac{\|\nabla D_q(\mathbf{y}_{(p)})\|}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|}=\sqrt{\lim_{\mathbf{y}_{(p)}\to\mathbf{x}_k}\frac{\|\nabla D_q(\mathbf{y}_{(p)})\|^2}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^2}}\leqslant\sqrt{\frac{\vartheta_J}{2}}. \tag{65}$$

Let $\zeta\triangleq\sqrt{\frac{\vartheta_J}{2}}$ and the proof is finished. □

B.10 Proof of Theorem 13

Proof. Since $\mathbf{y}_{(p)}\to\mathbf{x}_k$ and the data points are distinct, we can assume that $\mathbf{y}_{(p)}\notin\{\mathbf{x}_i\}_{i=1}^m$, $\forall p\geqslant P$ for some sufficiently large P. Therefore, $\mathbf{y}_{(p+1)}=\mathbf{T}_1(\mathbf{y}_{(p)})$, $\forall p\geqslant P$. We begin with an important equation:

$$\mathbf{y}_{(p+1)}-\mathbf{x}_k=\frac{\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2}(\mathbf{x}_i-\mathbf{x}_k)}{\eta_k^q\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{q-2}+\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2}}. \tag{66}$$

The key technique is to eliminate the singular term $\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{q-2}$ in the denominator of the right side of (66) and construct the rate of convergence simultaneously.

In the $q=1$ case, we divide both sides of (66) by the nonzero scalar $\|\mathbf{y}_{(p)}-\mathbf{x}_k\|$:

$$\frac{\mathbf{y}_{(p+1)}-\mathbf{x}_k}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|}=\frac{\sum_{i\neq k}\eta_i\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{-1}(\mathbf{x}_i-\mathbf{x}_k)}{\eta_k+\|\mathbf{y}_{(p)}-\mathbf{x}_k\|\cdot(\sum_{i\neq k}\eta_i\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{-1})}. \tag{67}$$

Taking L_2-norm $\|\cdot\|$ on both sides of (67) and letting $\mathbf{y}_{(p)}\to\mathbf{x}_k$ lead to

$$\begin{aligned}\lim_{\mathbf{y}_{(p)}\to\mathbf{x}_k}\frac{\|\mathbf{y}_{(p+1)}-\mathbf{x}_k\|}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|}&=\lim_{\mathbf{y}_{(p)}\to\mathbf{x}_k}\frac{\|\sum_{i\neq k}\eta_i\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{-1}(\mathbf{x}_i-\mathbf{x}_k)\|}{\eta_k+\|\mathbf{y}_{(p)}-\mathbf{x}_k\|\cdot(\sum_{i\neq k}\eta_i\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{-1})}\\&=\frac{\|-\nabla D_1(\mathbf{x}_k)\|}{\eta_k+0\cdot(\sum_{i\neq k}\eta_i\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{-1})}=\frac{\|\nabla D_1(\mathbf{x}_k)\|}{\eta_k}.\end{aligned} \tag{68}$$

Since $\eta_k>0$, the above limit is well-defined. Based on Theorem 6, the convergence is sublinear, linear or superlinear when $\|\nabla D_1(\mathbf{x}_k)\|=\eta_k$, $0<\|\nabla D_1(\mathbf{x}_k)\|<\eta_k$ or $\|\nabla D_1(\mathbf{x}_k)\|=0$, respectively.

In the $1<q<2$ case, it is a little subtle and we take two steps to eliminate the singular term $\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{q-2}$ in the denominator of (66). First, we divide both sides of (66) by the nonzero scalar $\|\mathbf{y}_{(p)}-\mathbf{x}_k\|$:

$$\frac{\mathbf{y}_{(p+1)}-\mathbf{x}_k}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|}=\frac{\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2}(\mathbf{x}_i-\mathbf{x}_k)}{\eta_k^q\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{q-1}+\|\mathbf{y}_{(p)}-\mathbf{x}_k\|\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})}. \tag{69}$$

Second, we multiply both the numerator and the denominator of the right side of (69) by $\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{1-q}$:

$$\begin{aligned}&\frac{\mathbf{y}_{(p+1)}-\mathbf{x}_k}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|}\\&=\frac{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{1-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2}(\mathbf{x}_i-\mathbf{x}_k))}{\eta_k^q+\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})}\\&=\frac{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{1-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})(\mathbf{y}_{(p)}-\mathbf{x}_k)}{\eta_k^q+\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})}-\frac{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{1-q}\cdot\nabla D_q(\mathbf{y}_{(p)})/q}{\eta_k^q+\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})}.\end{aligned} \tag{70}$$

Taking L_2-norm $\|\cdot\|$ on the leftmost side and the rightmost side of (70) leads to

$$\begin{aligned}
&\frac{\|\mathbf{y}_{(p+1)}-\mathbf{x}_k\|}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|}\\
\leqslant&\frac{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})}{\eta_k^q+\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})}+\frac{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{1-q}\cdot\|\nabla D_q(\mathbf{y}_{(p)})\|/q}{\eta_k^q+\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})}\\
=&\frac{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})}{\eta_k^q+\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})}+\frac{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}\cdot\left(\frac{\|\nabla D_q(\mathbf{y}_{(p)})\|}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|}\right)/q}{\eta_k^q+\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{y}_{(p)}-\mathbf{x}_i\|^{q-2})}. \qquad (71)
\end{aligned}$$

Because $1<q<2$, $0<2-q<1$ and $\|\mathbf{y}_{(p)}-\mathbf{x}_k\|^{2-q}$ is nonsingular when $\mathbf{y}_{(p)}\to\mathbf{x}_k$. Then we can adopt Lemma 12 to dominate the rightmost side of (71):

$$\lim_{\mathbf{y}_{(p)}\to\mathbf{x}_k}\frac{\|\mathbf{y}_{(p+1)}-\mathbf{x}_k\|}{\|\mathbf{y}_{(p)}-\mathbf{x}_k\|}\leqslant\frac{0\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{x}_k-\mathbf{x}_i\|^{q-2})+0\cdot\zeta/q}{\eta_k^q+0\cdot(\sum_{i\neq k}\eta_i^q\|\mathbf{x}_k-\mathbf{x}_i\|^{q-2})}=0, \qquad (72)$$

which shows a superlinear convergence for $1<q<2$. $\square$