Proof: See [YUAN2010]. ∎

- Note that the previous result for the gradient method Theorem 1.5.5 was only a local result.

- Comparing the rate of convergence of the gradient method for the classes $\mathcal{F}_L^{1,1}(\mathbb{R}^n)$ and $\mathcal{S}_{\mu,L}^{1,1}(\mathbb{R}^n)$, Theorems 2.5.1 (Corollary 2.5.2) and 2.5.3 with their lower complexity bounds, Theorems 2.2.1 and 2.4.1, respectively, we possible have a huge gap.

2.6 The optimal gradient method

Definition 2.6.1 A pair of sequences $\{\phi_k(\boldsymbol{x})\}_{k=0}^{\infty}$ and $\{\lambda_k\}_{k=0}^{\infty}$ with $\lambda_k \geq 0$ is called an *estimate sequence* of the function $f(\boldsymbol{x})$ if

$$\lambda_k \to 0,$$

and for any $\boldsymbol{x} \in \mathbb{R}^n$ and any $k \geq 0$, we have

$$\phi_k(\boldsymbol{x}) \leq (1-\lambda_k)f(\boldsymbol{x}) + \lambda_k \phi_0(\boldsymbol{x}).$$

Lemma 2.6.2 Given an estimate sequence $\{\phi_k(\boldsymbol{x})\}_{k=0}^{\infty}$, $\{\lambda_k\}_{k=0}^{\infty}$, and if for some sequence $\{\boldsymbol{x}_k\}_{k=1}^{\infty}$ we have

$$f(\boldsymbol{x}_k) \leq \phi_k^* \equiv \min_{\boldsymbol{x}\in\mathbb{R}^n} \phi_k(\boldsymbol{x})$$

then $f(\boldsymbol{x}_k) - f^* \leq \lambda_k(\phi_0(\boldsymbol{x}^*) - f(\boldsymbol{x}^*)) \to 0$.

Proof: It follows from the definition. ∎

Lemma 2.6.3 Assume that

1. $f \in \mathcal{S}_\mu^1(\mathbb{R}^n)$, possible with $\mu = 0$ (which means that $f \in \mathcal{F}^1(\mathbb{R}^n)$).

2. $\phi_0(\boldsymbol{x})$ is an arbitrary function on $\mathbb{R}^n$.

3. $\{\boldsymbol{y}_k\}_{k=0}^{\infty}$ is an arbitrary sequence in $\mathbb{R}^n$.

4. $\{\alpha_k\}_{k=0}^{\infty}$ is an arbitrary sequence such that $\alpha_k \in (0,1)$, $\sum_{k=0}^{\infty} \alpha_k = \infty$, and $\alpha_{-1} = 0$.

Then the pair of sequences $\{\Pi_{i=-1}^{k-1}(1-\alpha_i)\}_{k=0}^{\infty}$ and $\{\phi_k(\boldsymbol{x})\}_{k=0}^{\infty}$ recursively defined as

$$\phi_{k+1}(\boldsymbol{x}) \;=\; (1-\alpha_k)\phi_k(\boldsymbol{x}) + \alpha_k\left[f(\boldsymbol{y}_k) + \langle f'(\boldsymbol{y}_k), \boldsymbol{x}-\boldsymbol{y}_k\rangle + \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}_k\|^2\right]$$

is an estimate sequence.

Proof: Let us prove by induction on k. For $k = 0$, $\phi_0(\boldsymbol{x}) = (1-(1-\alpha_{-1}))\, f(\boldsymbol{x}) + (1-\alpha_{-1})\phi_0(\boldsymbol{x})$ since $\alpha_{-1} = 1$. Suppose that the induction hypothesis is valid for k. Since $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$,

$$\begin{aligned}
\phi_{k+1}(\boldsymbol{x}) &= (1-\alpha_k)\phi_k(\boldsymbol{x}) + \alpha_k \left[f(\boldsymbol{y}_k) + \langle f'(\boldsymbol{y}_k), \boldsymbol{x}-\boldsymbol{y}_k\rangle + \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}_k\|^2 \right] \\
&\leq (1-\alpha_k)\phi_k(\boldsymbol{x}) + \alpha_k f(\boldsymbol{x}) \\
&= \left(1-(1-\alpha_k)\Pi_{i=-1}^{k-1}(1-\alpha_i)\right) f(\boldsymbol{x}) + (1-\alpha_k)\left(\phi_k(\boldsymbol{x}) - (1-\Pi_{i=-1}^{k-1}(1-\alpha_i))f(\boldsymbol{x})\right) \\
&\leq \left(1-(1-\alpha_k)\Pi_{i=-1}^{k-1}(1-\alpha_i)\right) f(\boldsymbol{x}) + (1-\alpha_k)\Pi_{i=-1}^{k-1}(1-\alpha_i)\phi_0(\boldsymbol{x}) \\
&= \left(1-\Pi_{i=-1}^{k}(1-\alpha_i)\right) f(\boldsymbol{x}) + \Pi_{i=-1}^{k}(1-\alpha_i)\phi_0(\boldsymbol{x}).
\end{aligned}$$

The remaining part is left for exercise. ∎

Lemma 2.6.4 Let $\gamma_0, \phi_0^* \in \mathbb{R}$, $\mu \in \mathbb{R}$ (possible with $\mu = 0$), $\boldsymbol{v}_0 \in \mathbb{R}^n$, and $\{\boldsymbol{y}_k\}_{k=0}^\infty$ a given arbitrarily sequence. Define $\phi_0(\boldsymbol{x}) = \phi_0^* + \frac{\gamma_0}{2}\|\boldsymbol{x}-\boldsymbol{v}_0\|^2$. If we define recursively $\phi_{k+1}(\boldsymbol{x})$ such as the previous lemma:

$$\phi_{k+1}(\boldsymbol{x}) = (1-\alpha_k)\phi_k(\boldsymbol{x}) + \alpha_k \left[f(\boldsymbol{y}_k) + \langle f'(\boldsymbol{y}_k), \boldsymbol{x}-\boldsymbol{y}_k\rangle + \frac{\mu}{2}\|\boldsymbol{x}-\boldsymbol{y}_k\|^2 \right],$$

for an arbitrary sequence $\{\alpha_k\}_{k=0}^\infty$ such that $\alpha_k \in (0,1)$ and $\sum_{k=0}^\infty \alpha_k = \infty$. Then $\phi_{k+1}(\boldsymbol{x})$ preserve the canonical form

$$\phi_{k+1}(\boldsymbol{x}) = \phi_{k+1}^* + \frac{\gamma_{k+1}}{2}\|\boldsymbol{x}-\boldsymbol{v}_{k+1}\|^2 \tag{2.3}$$

for

$$\begin{aligned}
\gamma_{k+1} &= (1-\alpha_k)\gamma_k + \alpha_k\mu, \\
\boldsymbol{v}_{k+1} &= \frac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k\boldsymbol{v}_k + \alpha_k\mu\boldsymbol{y}_k - \alpha_k f'(\boldsymbol{y}_k)], \\
\phi_{k+1}^* &= (1-\alpha_k)\phi_k^* + \alpha_k f(\boldsymbol{y}_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|f'(\boldsymbol{y}_k)\|^2 \\
&\quad + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|\boldsymbol{y}_k-\boldsymbol{v}_k\|^2 + \langle f'(\boldsymbol{y}_k), \boldsymbol{v}_k-\boldsymbol{y}_k\rangle\right).
\end{aligned}$$

Proof: We will use again the induction hypothesis in k. Note that $\phi_0''(\boldsymbol{x}) = \gamma_0\boldsymbol{I}$. Now, for any $k \geq 0$,

$$\phi_{k+1}''(\boldsymbol{x}) = (1-\alpha_k)\phi_k''(\boldsymbol{x}) + \alpha_k\mu\boldsymbol{I} = ((1-\alpha_k)\gamma_k + \alpha_k\mu)\,\boldsymbol{I} = \gamma_{k+1}\boldsymbol{I}.$$

Therefore, $\phi_{k+1}(\boldsymbol{x})$ is a quadratic function of the form (2.3). From the first-order optimality condition

$$\begin{aligned}
\phi_{k+1}'(\boldsymbol{x}) &= (1-\alpha_k)\phi_k'(\boldsymbol{x}) + \alpha_k f'(\boldsymbol{y}_k) + \alpha_k\mu(\boldsymbol{x}-\boldsymbol{y}_k) \\
&= (1-\alpha_k)\gamma_k(\boldsymbol{x}-\boldsymbol{v}_k) + \alpha_k f'(\boldsymbol{y}_k) + \alpha_k\mu(\boldsymbol{x}-\boldsymbol{y}_k) = 0.
\end{aligned}$$

Thus,

$$\boldsymbol{x} = \boldsymbol{v}_{k+1} = \frac{1}{\gamma_{k+1}}\left[(1-\alpha_k)\gamma_k \boldsymbol{v}_k + \alpha_k \mu \boldsymbol{y}_k - \alpha_k f'(\boldsymbol{y}_k)\right]$$

is the minimal optimal solution of $\phi_{k+1}(\boldsymbol{x})$.

Finally, from what we proved so far and from the definition

$$\begin{array}{rcl} \phi_{k+1}(\boldsymbol{y}_k) & = & \phi^*_{k+1} + \frac{\gamma_{k+1}}{2}\|\boldsymbol{y}_k - \boldsymbol{v}_{k+1}\|^2 \\ & = & (1-\alpha_k)\phi_k(\boldsymbol{y}_k) + \alpha_k f(\boldsymbol{y}_k) \\ & = & (1-\alpha_k)\left(\phi^*_k + \frac{\gamma_k}{2}\|\boldsymbol{y}_k - \boldsymbol{v}_k\|^2\right) + \alpha_k f(\boldsymbol{y}_k). \end{array} \tag{2.4}$$

Now,

$$\boldsymbol{v}_{k+1} - \boldsymbol{y}_k = \frac{1}{\gamma_{k+1}}\left[(1-\alpha_k)\gamma_k(\boldsymbol{v}_k - \boldsymbol{y}_k) - \alpha_k f'(\boldsymbol{y}_k)\right].$$

Therefore,

$$\begin{array}{rcl} \frac{\gamma_{k+1}}{2}\|\boldsymbol{v}_{k+1} - \boldsymbol{y}_k\|^2 & = & \frac{1}{2\gamma_{k+1}}\left[(1-\alpha_k)^2\gamma_k^2\|\boldsymbol{v}_k - \boldsymbol{y}_k\|^2 + \alpha_k^2\|f'(\boldsymbol{y}_k)\|^2\right. \\ & & \left. -2\alpha_k(1-\alpha_k)\gamma_k\langle f'(\boldsymbol{y}_k), \boldsymbol{v}_k - \boldsymbol{y}_k\rangle\right]. \end{array} \tag{2.5}$$

Substituting (2.5) into (2.4), we obtain the expression for ϕ^*_{k+1}. ∎

Theorem 2.6.5 Consider $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, possible with $\mu = 0$ (which means that $f \in \mathcal{F}^{1,1}_L(\mathbb{R}^n)$). For a given $\boldsymbol{x}_0, \boldsymbol{v}_0 \in \mathbb{R}^n$ and $L \geq \gamma_0 \geq \mu \geq 0$, let us choose $\phi^*_0 = f(\boldsymbol{x}_0)$. Define the sequences $\{\alpha_k\}_{k=0}^\infty$, $\{\gamma_k\}_{k=0}^\infty$, $\{\boldsymbol{y}_k\}_{k=0}^\infty$, $\{\boldsymbol{x}_k\}_{k=0}^\infty$, $\{\boldsymbol{v}_k\}_{k=0}^\infty$, $\{\phi^*_k\}_{k=0}^\infty$, and $\{\phi_k\}_{k=0}^\infty$ as follows:

$$\begin{array}{rl} \alpha_k \in (0,1) \quad \text{root of} & L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu \equiv \gamma_{k+1}, \\ \boldsymbol{y}_k = & \dfrac{\alpha_k\gamma_k\boldsymbol{v}_k + \gamma_{k+1}\boldsymbol{x}_k}{\gamma_k + \alpha_k\mu}, \\ \boldsymbol{x}_k \quad \text{is such that} & f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{y}_k) - \dfrac{1}{2L}\|f'(\boldsymbol{y}_k)\|^2, \\ \boldsymbol{v}_{k+1} = & \dfrac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k\boldsymbol{v}_k + \alpha_k\mu\boldsymbol{y}_k - \alpha_k f'(\boldsymbol{y}_k)], \\ \phi^*_{k+1} = & (1-\alpha_k)\phi^*_k + \alpha_k f(\boldsymbol{y}_k) - \dfrac{\alpha_k^2}{2\gamma_{k+1}}\|f'(\boldsymbol{y}_k)\|^2 \\ & +\dfrac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\dfrac{\mu}{2}\|\boldsymbol{y}_k - \boldsymbol{v}_k\|^2 + \langle f'(\boldsymbol{y}_k), \boldsymbol{v}_k - \boldsymbol{y}_k\rangle\right), \\ \phi_{k+1}(\boldsymbol{x}) = & \phi^*_{k+1} + \dfrac{\gamma_{k+1}}{2}\|\boldsymbol{x} - \boldsymbol{v}_{k+1}\|^2. \end{array}$$

Then, we satisfy all the conditions of Lemma 2.6.2.

Proof: In fact, it just remains to show that $f(\boldsymbol{x}_k) \leq \phi^*_k$ and $\sum_{k=1}^\infty \alpha_k = \infty$.

For $k = 0$, $f(\boldsymbol{x}_0) \leq \phi^*_0$. Suppose that induction hypothesis is valid for k, and due to the previous lemma,

$$\phi^*_{k+1} \;=\; (1-\alpha_k)\phi^*_k + \alpha_k f(\boldsymbol{y}_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|f'(\boldsymbol{y}_k)\|^2$$