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

$$\begin{array}{rcl} \phi_{k+1}(\boldsymbol{y}_k) & = & \phi^*_{k+1} + \frac{\gamma_{k+1}}{2}\|\boldsymbol{y}_k - \boldsymbol{v}_{k+1}\|_2^2 \\ & = & (1-\alpha_k)\phi_k(\boldsymbol{y}_k) + \alpha_k f(\boldsymbol{y}_k) \\ & = & (1-\alpha_k)\left(\phi^*_k + \frac{\gamma_k}{2}\|\boldsymbol{y}_k - \boldsymbol{v}_k\|_2^2\right) + \alpha_k f(\boldsymbol{y}_k). \end{array} \tag{13}$$

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,

$$\begin{array}{rcl} \frac{\gamma_{k+1}}{2}\|\boldsymbol{v}_{k+1} - \boldsymbol{y}_k\|_2^2 & = & \frac{1}{2\gamma_{k+1}}\left[(1-\alpha_k)^2\gamma_k^2\|\boldsymbol{v}_k - \boldsymbol{y}_k\|_2^2 + \alpha_k^2\|f'(\boldsymbol{y}_k)\|_2^2\right. \\ & & \left.-2\alpha_k(1-\alpha_k)\gamma_k\langle f'(\boldsymbol{y}_k), \boldsymbol{v}_k - \boldsymbol{y}_k\rangle\right]. \end{array} \tag{14}$$

Substituting (14) into (13), we obtain the expression for ϕ^*_{k+1}. ∎

Theorem 9.5 Let $L \geq \mu \geq 0$. 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 given $\boldsymbol{x}_0, \boldsymbol{v}_0 \in \mathbb{R}^n$, let us choose $\phi^*_0 = f(\boldsymbol{x}_0)$. Consider also $\gamma_0 > 0$ such that $L \geq \gamma_0 \geq \mu \geq 0$. Define the sequences $\{\alpha_k\}_{k=-1}^\infty$, $\{\gamma_k\}_{k=0}^\infty$, $\{\boldsymbol{y}_k\}_{k=0}^\infty$, $\{\boldsymbol{x}_k\}_{k=0}^\infty$, $\{\boldsymbol{v}_k\}_{k=0}^\infty$, $\{\phi^*_k\}_{k=0}^\infty$, and $\{\phi_k(\boldsymbol{x})\}_{k=0}^\infty$ as follows:

$$\begin{array}{rcl} & \alpha_{-1} = 0, & \\ \alpha_k \in (0,1] & \text{root of} & L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu := \gamma_{k+1}, \\ & \boldsymbol{y}_k = & \dfrac{\alpha_k\gamma_k\boldsymbol{v}_k + \gamma_{k+1}\boldsymbol{x}_k}{\gamma_k + \alpha_k\mu}, \\ \boldsymbol{x}_k & \text{is such that} & f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{y}_k) - \dfrac{1}{2L}\|f'(\boldsymbol{y}_k)\|_2^2, \\ & \boldsymbol{v}_{k+1} = & \dfrac{1}{\gamma_{k+1}}[(1-\alpha_k)\gamma_k\boldsymbol{v}_k + \alpha_k\mu\boldsymbol{y}_k - \alpha_k f'(\boldsymbol{y}_k)], \\ & \phi^*_{k+1} = & (1-\alpha_k)\phi^*_k + \alpha_k f(\boldsymbol{y}_k) - \dfrac{\alpha_k^2}{2\gamma_{k+1}}\|f'(\boldsymbol{y}_k)\|_2^2 \\ & & +\dfrac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\dfrac{\mu}{2}\|\boldsymbol{y}_k - \boldsymbol{v}_k\|_2^2 + \langle f'(\boldsymbol{y}_k), \boldsymbol{v}_k - \boldsymbol{y}_k\rangle\right), \\ & \phi_{k+1}(\boldsymbol{x}) = & \phi^*_{k+1} + \dfrac{\gamma_{k+1}}{2}\|\boldsymbol{x} - \boldsymbol{v}_{k+1}\|_2^2. \end{array}$$

Then, we satisfy all the conditions of Lemma 9.2 for the $\lambda_k = \displaystyle\prod_{i=-1}^{k-1}(1-\alpha_i)$.

Proof:

In fact, due to Lemmas 9.3 and 9.4, it just remains to show that $\alpha_k \in (0,1]$ for $(k = 0, 1, \ldots)$ such that $\displaystyle\sum_{k=0}^\infty \alpha_k = \infty$. In the special case of $\mu = 0$, we must show that $\alpha_k < 1 \quad (k = 0, 1, \ldots)$. And finally that $f(\boldsymbol{x}_k) \leq \phi^*_k$.

Let us show both using induction hypothesis.

Consider the quadratic equation in α, $q_0(\alpha) := L\alpha^2 + (\gamma_0 - \mu)\alpha - \gamma_0 = 0$. Notice that its discriminant $\Delta := (\gamma_0 - \mu)^2 + 4\gamma_0 L$ is always positive by the hypothesis. Also, $q_0(0) = -\gamma_0 < 0$, but due to the hypothesis again. Therefore, this equation always has a root $\alpha_0 > 0$. Since $q_0(1) = L - \mu \geq 0$, $\alpha_0 \leq 1$, and we have $\alpha_0 \in (0,1]$. If $\mu = 0$, and $\alpha_0 = 1$, we will have $L = 0$ which implies $\gamma_0 = 0$ which contradicts our hypothesis. Then $\alpha_0 < 1$. In addition, $\gamma_1 := (1-\alpha_0)\gamma_0 + \alpha_0\mu > 0$ and $\gamma_0 + \alpha_0\mu > 0$. The same arguments are valid for any k. Therefore, $\alpha_k \in (0,1]$, and $\alpha_k < 1 \quad (k = 0, 1, \ldots,)$ if $\mu = 0$.

Finally, $L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu \geq (1-\alpha_k)\mu + \alpha_k\mu = \mu$. And we have $\alpha_k \geq \sqrt{\frac{\mu}{L}}$, and therefore, $\sum_{k=0}^{\infty} \alpha_k = \infty$, if $\mu > 0$. For the case $\mu = 0$, the argument is the same as the proof of Theorem 9.6.

Now, suppose that for $k = 0$, $f(\boldsymbol{x}_0) \leq \phi_0^*$. Suppose that the induction hypothesis is valid for any index equal or smaller than k. Due to the previous lemma,

$$\begin{aligned}
\phi_{k+1}^* &= (1-\alpha_k)\phi_k^* + \alpha_k f(\boldsymbol{y}_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|f'(\boldsymbol{y}_k)\|_2^2 \\
&\quad + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|\boldsymbol{y}_k - \boldsymbol{v}_k\|_2^2 + \langle f'(\boldsymbol{y}_k), \boldsymbol{v}_k - \boldsymbol{y}_k\rangle\right) \\
&\geq (1-\alpha_k)f(\boldsymbol{x}_k) + \alpha_k f(\boldsymbol{y}_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|f'(\boldsymbol{y}_k)\|_2^2 \\
&\quad + \frac{\alpha_k(1-\alpha_k)\gamma_k}{\gamma_{k+1}}\left(\frac{\mu}{2}\|\boldsymbol{y}_k - \boldsymbol{v}_k\|_2^2 + \langle f'(\boldsymbol{y}_k), \boldsymbol{v}_k - \boldsymbol{y}_k\rangle\right).
\end{aligned}$$

Now, since $f(\boldsymbol{x})$ is convex, $f(\boldsymbol{x}_k) \geq f(\boldsymbol{y}_k) + \langle f'(\boldsymbol{y}_k), \boldsymbol{x}_k - \boldsymbol{y}_k\rangle$, and we have:

$$\phi_{k+1}^* \geq f(\boldsymbol{y}_k) - \frac{\alpha_k^2}{2\gamma_{k+1}}\|f'(\boldsymbol{y}_k)\|_2^2 + (1-\alpha_k)\langle f'(\boldsymbol{y}_k), \frac{\alpha_k\gamma_k}{\gamma_{k+1}}(\boldsymbol{v}_k - \boldsymbol{y}_k) + \boldsymbol{x}_k - \boldsymbol{y}_k\rangle + \frac{\alpha_k(1-\alpha_k)\gamma_k\mu}{2\gamma_{k+1}}\|\boldsymbol{y}_k - \boldsymbol{v}_k\|_2^2.$$

Recall that since f' is L-Lipschitz continuous, if we apply Lemma 3.4 to $\boldsymbol{y}_k$ and $\boldsymbol{x}_{k+1} = \boldsymbol{y}_k - \frac{1}{L}f'(\boldsymbol{y}_k)$, we obtain

$$f(\boldsymbol{y}_k) - \frac{1}{2L}\|f'(\boldsymbol{y}_k)\|_2^2 \geq f(\boldsymbol{x}_{k+1}).$$

Therefore, if we impose

$$\frac{\alpha_k\gamma_k}{\gamma_{k+1}}(\boldsymbol{v}_k - \boldsymbol{y}_k) + \boldsymbol{x}_k - \boldsymbol{y}_k = \mathbf{0}$$

it justifies our choice for $\boldsymbol{y}_k$. And putting

$$\frac{\alpha_k^2}{2\gamma_{k+1}} = \frac{1}{2L}$$

it justifies our choice for α_k. Since $\frac{\alpha_k(1-\alpha_k)\gamma_k\mu}{\gamma_{k+1}} \geq 0$, we finally obtain $\phi_{k+1}^* \geq f(\boldsymbol{x}_{k+1})$ as wished. ∎

The above theorem suggests an algorithm to minimize $f \in \mathcal{S}_{\mu,L}^{1,1}(\mathbb{R}^n)$.

Notice that in the following optimal gradient method, we don't need the estimated sequence anymore.

General Scheme for the Optimal Gradient Method	
Step 0:	Choose $\boldsymbol{x}_0 \in \mathbb{R}^n$, let $\gamma_0 > 0$ such that $L \geq \gamma_0 \geq \mu \geq 0$. Set $\boldsymbol{v}_0 := \boldsymbol{x}_0$ and $k := 0$.
Step 1:	Compute $\alpha_k \in (0,1]$ from the equation $L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k\mu$.
Step 2:	Set $\gamma_{k+1} := (1-\alpha_k)\gamma_k + \alpha_k\mu$, $\boldsymbol{y}_k := \frac{\alpha_k\gamma_k\boldsymbol{v}_k + \gamma_{k+1}\boldsymbol{x}_k}{\gamma_k + \alpha_k\mu}$.
Step 3:	Compute $f(\boldsymbol{y}_k)$ and $f'(\boldsymbol{y}_k)$.
Step 4:	Find $\boldsymbol{x}_{k+1}$ such that $f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{y}_k) - \frac{1}{2L}\|f'(\boldsymbol{y}_k)\|_2^2$ using "line search".
Step 5:	Set $\boldsymbol{v}_{k+1} := \frac{(1-\alpha_k)\gamma_k\boldsymbol{v}_k + \alpha_k\mu\boldsymbol{y}_k - \alpha_k f'(\boldsymbol{y}_k)}{\gamma_{k+1}}$, $k := k+1$ and go to Step 1.

Theorem 9.6 Consider $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, possible with $\mu = 0$ (which means that $f \in \mathcal{F}^{1,1}_{L}(\mathbb{R}^n)$). The general scheme of the optimal gradient method generates a sequence $\{\boldsymbol{x}_k\}_{k=0}^{\infty}$ such that

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \leq \lambda_k \left[f(\boldsymbol{x}_0) + \frac{\gamma_0}{2}\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2^2 - f(\boldsymbol{x}^*) \right],$$

where $\alpha_{-1} = 0$ and $\lambda_k = \prod_{i=-1}^{k-1}(1 - \alpha_i)$. Moreover,

$$\lambda_k \leq \min \left\{ \left(1 - \sqrt{\frac{\mu}{L}}\right)^k, \frac{4L}{(2\sqrt{L} + k\sqrt{\gamma_0})^2} \right\}.$$

In other words, the sequence $\{f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)\}_{k=0}^{\infty}$ converges R-sublinearly to zero if $\mu = 0$ and R-linearly to zero if $\mu > 0$.

Proof:

The first part is obvious from the definition and Lemma 9.2.

We already know that $\alpha_k \geq \sqrt{\frac{\mu}{L}}$ $(k = 0, 1, \ldots)$, therefore,

$$\lambda_k = \prod_{i=-1}^{k-1}(1 - \alpha_i) = \prod_{i=0}^{k-1}(1 - \alpha_i) \leq \left(1 - \sqrt{\frac{\mu}{L}}\right)^k,$$

which only has a meaning if $\mu > 0$. For the case $\mu = 0$, let us prove first that $\gamma_k = \gamma_0\lambda_k$. Obviously $\gamma_0 = \gamma_0\lambda_0$, and assuming the induction hypothesis,

$$\gamma_{k+1} = (1 - \alpha_k)\gamma_k + \alpha_k\mu = (1 - \alpha_k)\gamma_k = (1 - \alpha_k)\gamma_0\lambda_k = \gamma_0\lambda_{k+1}.$$

Therefore, $L\alpha_k^2 = \gamma_{k+1} = \gamma_0\lambda_{k+1}$. Since λ_k is a decreasing sequence

$$\begin{aligned}
\frac{1}{\sqrt{\lambda_{k+1}}} - \frac{1}{\sqrt{\lambda_k}} &= \frac{\sqrt{\lambda_k} - \sqrt{\lambda_{k+1}}}{\sqrt{\lambda_k\lambda_{k+1}}} = \frac{\lambda_k - \lambda_{k+1}}{\sqrt{\lambda_k\lambda_{k+1}}(\sqrt{\lambda_k} + \sqrt{\lambda_{k+1}})} \\
&\geq \frac{\lambda_k - \lambda_{k+1}}{\sqrt{\lambda_k\lambda_{k+1}}(\sqrt{\lambda_k} + \sqrt{\lambda_k})} = \frac{\lambda_k - \lambda_{k+1}}{2\lambda_k\sqrt{\lambda_{k+1}}} = \frac{\lambda_k - (1 - \alpha_k)\lambda_k}{2\lambda_k\sqrt{\lambda_{k+1}}} \\
&= \frac{\alpha_k}{2\sqrt{\lambda_{k+1}}} = \frac{1}{2}\sqrt{\frac{\gamma_0}{L}}.
\end{aligned}$$

Thus

$$\frac{1}{\sqrt{\lambda_k}} \geq 1 + \frac{k}{2}\sqrt{\frac{\gamma_0}{L}}$$

and we have the result. ∎

Theorem 9.7 Consider $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, possible with $\mu = 0$ (which means that $f \in \mathcal{F}^{1,1}_{L}(\mathbb{R}^n)$). If we take $\gamma_0 = L$, the general scheme of the "optimal" gradient method generates a sequence $\{\boldsymbol{x}_k\}_{k=0}^{\infty}$ such that

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \leq L \min \left\{ \left(1 - \sqrt{\frac{\mu}{L}}\right)^k, \frac{4}{(k+2)^2} \right\} \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2.$$

This means that it is "optimal" for the class of functions from $\mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$ with $\mu > 0$, or $\mathcal{F}^{1,1}_{L}(\mathbb{R}^n)$.

In the particular case of $\mu > 0$, we have the following inequality for k sufficiently large:

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 \leq \frac{2L}{\mu}\left(1 - \sqrt{\frac{\mu}{L}}\right)^k \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2.$$

That means that the sequence $\{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2\}_{k=0}^{\infty}$ converges Q-lineary to zero.