Finally, if $\mu > 0$ and we choose $\gamma_0 := \alpha_0(\alpha_0 L - \mu)/(1-\alpha_0) = \mu$, we have a further simplification.

$$\alpha_k = \sqrt{\frac{\mu}{L}}, \quad \beta_k = \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}$$

and we end up with

Constant Step Scheme II for the "Optimal" Gradient Method	
Step 0:	Choose $\boldsymbol{x}_0 \in \mathbb{R}^n$, set $\boldsymbol{y}_0 := \boldsymbol{x}_0$ and $k := 0$.
Step 1:	Compute $f'(\boldsymbol{y}_k)$.
Step 2:	Set $\boldsymbol{x}_{k+1} := \boldsymbol{y}_k - \frac{1}{L} f'(\boldsymbol{y}_k)$.
Step 3:	Set $\boldsymbol{y}_{k+1} := \boldsymbol{x}_{k+1} + \frac{\sqrt{L}-\sqrt{\mu}}{\sqrt{L}+\sqrt{\mu}}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)$, $k := k+1$ and go to Step 1.

You can find a variation of this method for instance in: C. C. Gonzaga and E. W. Karas, "Fine tuning Nesterov's steepest descent algorithm for differentiable convex programming," *Mathematical Programming*, **138** (2013), pp. 141–166.

9.1 Exercises

1. Complete the proof of Lemma 9.3.

2. We want to justify the Constant Step Scheme I of the "Optimal" Gradient Method. This is a particular case of the General Scheme for the "Optimal" Gradient Method for the following choice:

$$\begin{aligned}
\gamma_{k+1} &:= L\alpha_k^2 = (1-\alpha_k)\gamma_k + \alpha_k \mu \\
\boldsymbol{y}_k &= \frac{\alpha_k \gamma_k \boldsymbol{v}_k + \gamma_{k+1}\boldsymbol{x}_k}{\gamma_k + \alpha_k \mu} \\
\boldsymbol{x}_{k+1} &= \boldsymbol{y}_k - \frac{1}{L} f'(\boldsymbol{y}_k) \\
\boldsymbol{v}_{k+1} &= \frac{(1-\alpha_k)\gamma_k \boldsymbol{v}_k + \alpha_k \mu \boldsymbol{y}_k - \alpha_k f'(\boldsymbol{y}_k)}{\gamma_{k+1}}.
\end{aligned}$$

   (a) Show that $\boldsymbol{v}_{k+1} = \boldsymbol{x}_k + \frac{1}{\alpha_k}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)$.

   (b) Show that $\boldsymbol{y}_{k+1} = \boldsymbol{x}_{k+1} + \beta_k(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)$ for $\beta_k = \frac{\alpha_{k+1}\gamma_{k+1}(1-\alpha_k)}{\alpha_k(\gamma_{k+1}+\alpha_{k+1}\mu)}$.

   (c) Show that $\beta_k = \frac{\alpha_k(1-\alpha_k)}{\alpha_k^2+\alpha_{k+1}}$.

   (d) Explain why $\alpha_{k+1}^2 = (1-\alpha_{k+1})\alpha_k^2 + \frac{\mu}{L}\alpha_{k+1}$.

10 Extension of the "Optimal" Gradient Method (Accelerated Gradient Method) for "Simple" Convex Sets

We are interested now to solve the following problem:

$$\left\{ \begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in Q \end{array} \right. \tag{15}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and Q is a closed convex subset of $\mathbb{R}^n$, simple enough to have an easy projection onto it, *e.g.*, positive orthant, n dimensional box, simplex, Euclidean ball, *etc.*

Lemma 10.1 Let $f \in \mathcal{F}^1(\mathbb{R}^n)$ and Q be a closed convex set. The point $\boldsymbol{x}^*$ is a solution of (15) if and only if

$$\langle f'(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle \geq 0, \quad \forall \boldsymbol{x} \in Q.$$

Proof:
Indeed, if the inequality is true,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \langle f'(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle \geq f(\boldsymbol{x}^*) \quad \forall \boldsymbol{x} \in Q.$$

For the converse, let $\boldsymbol{x}^*$ be an optimal solution of the minimization problem (15). Assume by contradiction that there is a $\boldsymbol{x} \in Q$ such that $\langle f'(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle < 0$. Consider the function $\phi(\alpha) = f(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*))$ for $\alpha \in [0, 1]$. Then, $\phi(0) = f(\boldsymbol{x}^*)$ and $\phi'(0) = \langle f'(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle < 0$. Therefore, for $\alpha > 0$ small enough, we have

$$f(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)) = \phi(\alpha) < \phi(0) = f(\boldsymbol{x}^*)$$

which is a contradiction. ∎

Lemma 10.2 Let $f \in \mathcal{S}^1_\mu(\mathbb{R}^n)$ with $\mu > 0$, and Q be a closed convex set. Then there exists a unique solution $\boldsymbol{x}^*$ for the problem (15).

Proof:
Left for exercise ∎

Definition 10.3 Let $f \in \mathcal{C}^1(\mathbb{R}^n)$, Q a closed convex set, $\bar{\boldsymbol{x}} \in \mathbb{R}^n$, and $\gamma > 0$. Denote by

$$\begin{aligned} \boldsymbol{x}_Q(\bar{\boldsymbol{x}}; \gamma) &:= \arg\min_{\boldsymbol{y} \in Q} \left[f(\bar{\boldsymbol{x}}) + \langle f'(\bar{\boldsymbol{x}}), \boldsymbol{y} - \bar{\boldsymbol{x}} \rangle + \frac{\gamma}{2} \|\boldsymbol{y} - \bar{\boldsymbol{x}}\|_2^2 \right], \\ \boldsymbol{g}_Q(\bar{\boldsymbol{x}}; \gamma) &:= \gamma(\bar{\boldsymbol{x}} - \boldsymbol{x}_Q(\bar{\boldsymbol{x}}; \gamma)). \end{aligned}$$

We call $\boldsymbol{g}_Q(\bar{\boldsymbol{x}}; \gamma)$ the *gradient mapping of f on Q*. Observe that due to Lemma 10.2, $\boldsymbol{x}_Q(\bar{\boldsymbol{x}}; \gamma)$ exists and it is uniquely defined.

In the case $Q \equiv \mathbb{R}^n$, notice that $\boldsymbol{x}_Q(\bar{\boldsymbol{x}}; \gamma) = \bar{\boldsymbol{x}} - \frac{1}{\gamma} f'(\bar{\boldsymbol{x}})$ and $\boldsymbol{g}_Q(\bar{\boldsymbol{x}}; \gamma) = f'(\bar{\boldsymbol{x}})$. Therefore, they take the roles of $\boldsymbol{x}_{k+1}$ and $\boldsymbol{y}_k$ in the Constant Step Scheme I for the "Optimal" Gradient Method.

Theorem 10.4 Let $f \in \mathcal{S}^{1,1}_{\mu,L}(\mathbb{R}^n)$, $\gamma \geq L$, $\gamma > 0$, Q a closed convex set, and $\bar{\boldsymbol{x}} \in \mathbb{R}^n$. Then

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}_Q(\bar{\boldsymbol{x}}; \gamma)) + \langle \boldsymbol{g}_Q(\bar{\boldsymbol{x}}; \gamma), \boldsymbol{x} - \bar{\boldsymbol{x}} \rangle + \frac{1}{2\gamma} \|\boldsymbol{g}_Q(\bar{\boldsymbol{x}}; \gamma)\|_2^2 + \frac{\mu}{2} \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2^2, \quad \forall \boldsymbol{x} \in Q.$$

Proof:
Let us use the following notation $\boldsymbol{x}_Q := \boldsymbol{x}_Q(\bar{\boldsymbol{x}}; \gamma)$ and $\boldsymbol{g}_Q := \boldsymbol{g}_Q(\bar{\boldsymbol{x}}; \gamma)$. Consider $\phi(\boldsymbol{x}) := f(\bar{\boldsymbol{x}}) + \langle f'(\bar{\boldsymbol{x}}), \boldsymbol{x} - \bar{\boldsymbol{x}} \rangle + \frac{\gamma}{2} \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2^2$.
Then $\phi'(\boldsymbol{x}) = f'(\bar{\boldsymbol{x}}) + \gamma(\boldsymbol{x} - \bar{\boldsymbol{x}})$. Therefore $\forall \boldsymbol{x} \in Q$, we have

$$\langle \phi'(\boldsymbol{x}_Q), \boldsymbol{x} - \boldsymbol{x}_Q \rangle = \langle f'(\bar{\boldsymbol{x}}) + \gamma(\boldsymbol{x}_Q - \bar{\boldsymbol{x}}), \boldsymbol{x} - \boldsymbol{x}_Q \rangle = \langle f'(\bar{\boldsymbol{x}}) - \boldsymbol{g}_Q, \boldsymbol{x} - \boldsymbol{x}_Q \rangle \geq 0$$

due to Lemma 10.1.
Hence, $\forall \boldsymbol{x} \in Q$,

$$\begin{aligned} f(\boldsymbol{x}) - \frac{\mu}{2} \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2^2 &\geq f(\bar{\boldsymbol{x}}) + \langle f'(\bar{\boldsymbol{x}}), \boldsymbol{x} - \bar{\boldsymbol{x}} \rangle \\ &= f(\bar{\boldsymbol{x}}) + \langle f'(\bar{\boldsymbol{x}}), \boldsymbol{x} - \boldsymbol{x}_Q \rangle + \langle f'(\bar{\boldsymbol{x}}), \boldsymbol{x}_Q - \bar{\boldsymbol{x}} \rangle \\ &\geq f(\bar{\boldsymbol{x}}) + \langle \boldsymbol{g}_Q, \boldsymbol{x} - \boldsymbol{x}_Q \rangle + \langle f'(\bar{\boldsymbol{x}}), \boldsymbol{x}_Q - \bar{\boldsymbol{x}} \rangle \\ &= \phi(\boldsymbol{x}_Q) - \frac{\gamma}{2} \|\boldsymbol{x}_Q - \bar{\boldsymbol{x}}\|_2^2 + \langle \boldsymbol{g}_Q, \boldsymbol{x} - \boldsymbol{x}_Q \rangle \end{aligned}$$