2.2.3 Gradient mapping

In the constrained minimization problem the gradient of the objective function should be treated differently as compared to the unconstrained situation. In the previous section we have already seen that its role in optimality conditions is changing. Moreover, we cannot use it anymore in a gradient step since the result could be infeasible, etc. If we look at the main properties of the gradient, which we have used for $f \in \mathcal{F}_L^{1,1}(R^n)$, we can see that two of them are of the most importance. The first one is that the gradient step decreases the function value by an amount comparable with the squared norm of the gradient:

$$f(x - \tfrac{1}{L}f'(x)) \le f(x) - \tfrac{1}{2L} \parallel f'(x) \parallel^2 .$$

And the second one is the inequality

$$\langle f'(x), x - x^* \rangle \ge \tfrac{1}{L} \parallel f'(x) \parallel^2 .$$

It turns out that for constrained minimization problems we can introduce an object that inherits the most important properties of the gradient.

DEFINITION 2.2.3 *Let us fix some $\gamma > 0$. Denote*

$$\begin{aligned} x_Q(\bar{x};\gamma) &= \arg\min_{x \in Q} \left[f(\bar{x}) + \langle f'(\bar{x}), x - \bar{x} \rangle + \tfrac{\gamma}{2} \parallel x - \bar{x} \parallel^2 \right], \\ g_Q(\bar{x};\gamma) &= \gamma(\bar{x} - x_Q(\bar{x};\gamma)). \end{aligned}$$

We call $g_Q(\gamma, x)$ the gradient mapping *of f on Q.*

For $Q \equiv R^n$ we have

$$x_Q(\bar{x};\gamma) = \bar{x} - \tfrac{1}{\gamma} f'(\bar{x}), \quad g_Q(\bar{x};\gamma) = f'(\bar{x}).$$

Thus, the value $\frac{1}{\gamma}$ can be seen as a step size for the "gradient" step

$$\bar{x} \to x_Q(\bar{x};\gamma).$$

Note that the gradient mapping is well defined in view of Theorem 2.2.6. Moreover, it is defined for all $\bar{x} \in R^n$, not necessarily from Q.

Let us write down the main property of gradient mapping.

THEOREM 2.2.7 *Let $f \in \mathcal{S}_{\mu,L}^{1,1}(R^n)$, $\gamma \ge L$ and $\bar{x} \in R^n$. Then for any $x \in Q$ we have*

$$\begin{aligned} f(x) \ \ge \ & f(x_Q(\bar{x};\gamma)) + \langle g_Q(\bar{x};\gamma), x - \bar{x} \rangle \\ & + \tfrac{1}{2\gamma} \parallel g_Q(\bar{x};\gamma) \parallel^2 + \tfrac{\mu}{2} \parallel x - \bar{x} \parallel^2 . \end{aligned} \qquad (2.2.15)$$

Proof: Denote $x_Q = x_Q(\gamma, \bar{x})$, $g_Q = g_Q(\gamma, \bar{x})$ and let

$$\phi(x) = f(\bar{x}) + \langle f'(\bar{x}), x - \bar{x}\rangle + \frac{\gamma}{2} \| x - \bar{x} \|^2 .$$

Then $\phi'(x) = f'(\bar{x}) + \gamma(x - \bar{x})$, and for any $x \in Q$ we have

$$\langle f'(\bar{x}) - g_Q, x - x_Q\rangle = \langle \phi'(x_Q), x - x_Q\rangle \geq 0.$$

Hence,

$$\begin{aligned}
f(x) - \tfrac{\mu}{2} \| x - \bar{x} \|^2 &\geq f(\bar{x}) + \langle f'(\bar{x}), x - \bar{x}\rangle \\
&= f(\bar{x}) + \langle f'(\bar{x}), x_Q - \bar{x}\rangle + \langle f'(\bar{x}), x - x_Q\rangle \\
&\geq f(\bar{x}) + \langle f'(\bar{x}), x_Q - \bar{x}\rangle + \langle g_Q, x - x_Q\rangle \\
&= \phi(x_Q) - \tfrac{\gamma}{2} \| x_Q - \bar{x} \|^2 + \langle g_Q, x - x_Q\rangle \\
&= \phi(x_Q) - \tfrac{1}{2\gamma} \| g_Q \|^2 + \langle g_Q, x - x_Q\rangle \\
&= \phi(x_Q) + \tfrac{1}{2\gamma} \| g_Q \|^2 + \langle g_Q, x - \bar{x}\rangle,
\end{aligned}$$

and $\phi(x_Q) \geq f(x_Q)$ since $\gamma \geq L$. □

COROLLARY 2.2.1 *Let $f \in \mathcal{S}^{1,1}_{\mu,L}(R^n)$, $\gamma \geq L$ and $\bar{x} \in R^n$. Then*

$$f(x_Q(\bar{x};\gamma)) \leq f(\bar{x}) - \frac{1}{2\gamma} \| g_Q(\bar{x};\gamma) \|^2, \qquad (2.2.16)$$

$$\langle g_Q(\bar{x};\gamma), \bar{x} - x^*\rangle \geq \frac{1}{2\gamma} \| g_Q(\bar{x};\gamma) \|^2 + \frac{\mu}{2} \| x - \bar{x} \|^2 . \qquad (2.2.17)$$

Proof: Indeed, using (2.2.15) with $x = \bar{x}$, we get (2.2.16). Using (2.2.15) with $x = x^*$, we get (2.2.17) since $f(x_Q(\bar{x};\gamma)) \geq f(x^*)$. □

2.2.4 Minimization methods for simple sets

Let us show how we can use the gradient mapping for solving the following problem:

$$\min_{x \in Q} f(x),$$